Genetic Characterization and High Resolution Analysis of *Lr75*: a Novel Adult Plant Leaf Rust Resistance Gene in Wheat

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Jyoti Singla

aus

Indien

Promotionskomitee

Prof. Dr. Beat Keller (Vorsitz und Leitung der Dissertation)

PD Dr. Thomas Wicker

Prof. Dr. Ueli Grossnikalaus

Zürich, 2016

Table of Contents

Table of Contents I
Summary V
Zusammenfassung VII
Chapter 1 1
1.1 Importance of wheat in agriculture, its evolution and domestication 2
1.2 Diseases of wheat and wheat leaf rust 3
1.3 Leaf rust (*Puccinia triticina*) - an obligate biotrophic wheat fungal pathogen 3
1.3.1 Economic importance of wheat leaf rust 5
1.4 Disease resistance in crop plants 5
1.5 Genetic mapping and map-based cloning in wheat: Concepts and challenges 7
1.6 Technical advances to facilitate map-based cloning in wheat 8
1.6.1 Use of syntenic information between wheat and related model grass species 8
1.6.2 Advances in wheat genome sequencing 9
1.6.3 Advancements in molecular marker techniques 10
1.6.4 Applications of NGS in wheat 11
1.7 Aim of the thesis 12
Chapter 2 13
2.1 Summary 14
2.2 Introduction 15
2.3 Materials and Methods 17
2.3.1 Plant material 17
2.3.2 Characterization of leaf rust resistance in the field 17
2.3.3 Characterization of leaf rust resistance in the greenhouse 17
2.3.4 Measurement of pustule density of urediniospores 18
2.3.5 Marker analysis and genetic linkage mapping 18
2.3.6 Deletion bin mapping 19
2.4 Results 20
2.4.1 Selection of 'Forno' leaf rust QTLs for backcrossing 20
2.4.2 Development and evaluation of a near isogenic line, 'Arina*LrFor*', carrying *Lr75* and *QLr.sfr7BL* 21
2.4.3 'Arina*LrFor*' shows race-specific resistance at seedling stage 24
2.4.4 Genetic mapping of *Lr75* 25
2.4.5 Deletion bin mapping 26
2.5 Discussion 28
2.5.1 CIMMYT wheat germplasm as the major source of slow-rusting gene 28
2.5.2 Breeding for slow-rusting resistance in European wheat germplasm 29
2.5.3 Slow-rusting APR genes are influenced by environment 30

2.5.4 *QLr.sfr-7BL* might be the leaf rust resistance gene *Lr14a* 31
Chapter 3 33
3.1 Summary 34
3.2 Introduction 35
3.3 Materials and Methods 37
3.3.1 Plant material 37
3.3.2 Flow-sorting of chromosome 1B 37
3.3.3 Sequencing and assembly of chromosome 1B of 'Arina' and 'Forno' 37
3.3.4 Identification of SNPs for genotyping 37
3.3.5 Genetic map construction 37
3.3.6 Marker development using high resolution melting (HRM) analysis 38
3.4 Results 39
3.4.1 Use of flow cytometry sorting to generate a chromosome specific SNP map 39
3.4.1.1 Purification and sequencing of chromosome 1B of 'Arina' and 'Forno' 39
3.4.1.2 Gene identification using *Brachypodium* syntenic information and SNP detection in gene-containing 'Forno' contigs 39
3.4.1.3 Development of a genetic map of chromosome 1B based on SNP markers 42
3.4.2 90K SNP array as a tool for marker development 44
3.4.3 Use of BAC end sequences (BES) of wheat chromosome 1BS to design one new proximal flanking marker 45
3.4.4 Collinearity of the marker order between genetic maps of the RIL and NIL populations 47
3.4.5 Screening of wheat germplasm with a subset of developed SNP markers to determine the origin of *Lr75* 47
3.5 Discussion 50
3.5.1 High specificity of gene-associated SNP markers derived from flow-sorted chromosome sequences 50
3.5.2 90K SNP array as a tool for high resolution mapping 52
3.5.3 Advantages of different approaches used in this study for marker development 52
Chapter 4 54
4.1 Summary 55
4.2 Introduction 56
4.3 Material and Methods 58
4.3.1 Plant material 58
4.3.2 Screening of the cv. 'Chinese Spring' BAC library and BAC DNA isolation 58
4.3.3 BAC clone sequencing and sequence assembly of cv. 'Chinese Spring' 58
4.3.4 Orientation and positioning of assembled sequence contigs 58
4.3.5 Development of amplification products for screening the 'Forno' BAC library 59
4.3.6 Construction of non-gridded BAC library from cv. 'Forno' 60

4.3.7 BAC sequencing and assembly 60
4.3.8 Gene annotation 61
4.4 Results 62
4.4.1 Identification of a physical region spanning the *Lr75* locus using the 1BS physical map of wheat cv. 'Chinese Spring' 62
4.4.2 Sequencing of MTP BAC clones in the 3.0 Mb target region 62
4.4.3 SNP development from the flow-sorted Illumina sequence contigs of 'Arina' and 'Forno' to map the 24 BAC clones on the genetic map 66
4.4.4 Sequence characterization of the 723 kb physical target region of cv. 'Chinese Spring' 74
4.4.5 Comparison of the genetic and physical maps 75
4.4.6 BAC clones spanning the *Lr75* region in leaf rust resistant cv. 'Forno' 75
4.4.7 Gene annotation of the cv. 'Chinese Spring' and cv. 'Forno' target region for candidate gene identification 76
4.4.8 Haplotype comparison between 'Chinese Spring' and 'Forno' physical target intervals 78
4.5 Discussion 81
4.5.1 LTC assembly of BAC contigs improves quality of physical mapping 81
4.5.2 The leaf rust resistance gene *Lr75* might be different from the class of NBS-LRR resistance genes 81
4.5.3 Sequence divergence between haplotypes 82
4.5.4 Candidate genes identified in the 'Forno' physical target region 83
4.5.5 Evolution of chimeric genes 84
4.6 Outlook 85
Chapter 5 86
5.1 Race-specificity of 'Arina*LrFor*' at the seedling stage 87
5.1.1 Suppression of resistance by the presence of suppressor genes 87
5.1.2 Putative interaction of *Lr75* or *QLr.sfr-7BL* with a seedling resistance gene in the 'Arina' background 88
5.2 Origin of *Lr75* in the wheat germplasm 89
5.3 High-throughput genotyping in wheat 89
5.3.1 Exome capture as an efficient tool for targeted-capture resequencing of parts of wheat genome 90
5.3.2 SNP genotyping in wheat 91
5.4 Cloning of *Lr75* and approaches for validation of candidate genes 92
References 93
Appendix A 107
Appendix B 108
Appendix C 109
Appendix D 111
Appendix E 112

Appendix F 113
Appendix G 114
Appendix H 116
Appendix I 117
Appendix J 118
Acronyms 119
Acknowledgements 120
Curriculum vitae 122

Summary

Leaf rust is a devastating fungal disease of wheat. To date, only few partial adult plant leaf rust resistance genes have been described molecularly. The identification of more durable sources of leaf rust resistance as well as the combination of additively acting, partial adult plant leaf rust resistance genes has been a suitable strategy for breeding rust resistant wheat cultivars. This thesis describes the 1) marker-assisted introgression of two partial resistance genes, *QLr.sfr-1BS* and *QLr.sfr-7BL* to achieve broad-spectrum leaf rust resistance in the Swiss winter wheat cv. 'Arina' and the 2) genetic characterization, high resolution genetic and physical mapping of the partial adult plant leaf rust resistance gene, *QLr.sfr-1BS* present in the Swiss winter wheat cv. 'Forno'. Near-immunity to leaf rust in 'Forno' is conferred by the presence of at least six quantitative trait loci (QTL). By using a marker-based strategy we introgressed two additive leaf rust resistance QTLs without the undesired phenotypic trait leaf tip necrosis (LTN) from 'Forno' into the leaf rust susceptible Swiss winter wheat cv. 'Arina'. The resulting backcross line, 'Arina*LrFor*' showed improved leaf rust resistance when tested in multiple locations over several years. One of the introgressed QTL, *QLr.sfr-1BS* had previously been located to a 16 cM interval on chromosome 1BS of cv. 'Forno'. Because no other described leaf rust resistance gene has been identified in the same genetic region on chromosome 1BS, *QLr.sfr-1BS* is a novel gene and was designated *Lr75*.

In order to further saturate the *Lr75* target region with molecular markers, we used four different approaches for marker development: 1) the Illumina survey sequences of wheat cv. 'Chinese Spring' to generate chromosome 1BS-specific microsatellite markers, 2) sequenced flow-sorted chromosome 1B of the wheat cv. 'Arina' and 'Forno', 3) the wheat 90K SNP array and 4) the BAC end sequences (BES) of chromosome 1BS of the hexaploid wheat cv. 'Chinese Spring'. Using all the four approaches, a total of 98 microsatellite primers and 149 SNP-based primers were developed. Out of these, six microsatellite and 68 SNP markers were mapped in the 65 BC_3F_2 lines of a near isogenic line (NIL) population. The high resolution genetic mapping of newly developed microsatellite and SNP markers resulted in narrowing down of the *Lr75* target interval from 16 cM to 1.2 cM flanked by markers *B2g38480* and *B2g38540a*.

The corresponding 3 Mb physical target interval was identified by comparing the flanking marker sequences against the chromosome 1BS-specific BES of the fingerprinted BAC clones obtained by both fingerprinted contig (FPC) and linear topological contig (LTC) assembly of 'Chinese Spring'. The 24 BAC clones spanning the whole 3 Mb physical interval were completely sequenced and mapped on the NIL genetic map by development of gene-associated SNP markers from the BAC clones. Out of 24 newly developed SNP markers, 15 were polymorphic and their mapping resulted in the establishment of a further, smaller 0.96 cM target interval which corresponded to 723 kb in

‘Chinese Spring’. In addition, a yet incomplete 403.6 kb physical *Lr75* target interval from the *Lr75* donor cv. ‘Forno’ was established by constructing a ‘Forno’ non-gridded BAC library having one-fold coverage. The physical target interval in ‘Forno’ contains one gap of unknown size which needs to be covered by screening the ‘Forno’ BAC library with new primers. The manual assembly and annotation of the *Lr75* physical target regions of both haplotypes identified 14 and eight putative candidate genes in ‘Chinese Spring’ and ‘Forno’, respectively. Comparison of the two haplotypes of ‘Chinese Spring’ and ‘Forno’ revealed a series of insertion, deletion and duplication events, thereby revealing a considerable divergence between two haplotypes.

From this study we conclude that 1) marker-assisted combination of partial adult plant resistance genes is a feasible strategy to increase broad-spectrum disease resistance in winter wheat and 2) the different strategies presented in this study enable a fast and efficient development of genome-specific, gene-associated SNP markers which can be used for high resolution genetic mapping and map-based cloning.

Zusammenfassung

Braunrost ist eine verheerende Pilzkrankheit in Weizen, für welche bis heute nur wenige partielle ‚adult plant' Braunrostresistenzgene molekular beschrieben sind. Die Identifikation weiterer dauerhafter Quellen von Braunrostresistenz als auch ihre Kombination ist eine geeignete Strategie für die Züchtung von krankeitsresistenten Weizenkultivaren. Diese Thesis beschreibt 1) die Marker-assistierte Introgression der zwei partiellen Resistenzgene, *QLr.sfr-1BS* und *QLr.sfr-7BL*, zur Erreichung einer Breitspektrums-Resistenz gegen Braunrost in der Schweizer Winterweizensorte ‚Arina' und 2) die genetische Charakterisierung und hochauflösende genetische und physische Kartierung des partiellen ‚adult plant' Braunrostresistenzgens, *QLr.sfr-1BS*, das in der Schweizer Winterweizensorte ‚Forno' vorhanden ist. Die Anwesenheit von mindestens sechs „Quantitative Trait Loci" (QTL) verleiht ‚Forno' beinahe komplette Immunität gegen Braunrost. Einer dieser QTL ist mit dem unerwünschten phänotypischen Merkmal, ‚Leaf Tip Necrosis' (LTN) assoziiert. Unter Verwendung von Marker-assistertier Selektion kreuzten wir zwei additive, mit LTN nicht assoziierte, Braunrostresistenz-QTL von ‚Forno' in die braunrostanfällige Sorte ‚Arina' ein. Die resultierende Rückkreuzungslinie, ‚Arina*LrFor*' zeigte an verschiedenen Standorten und über mehrere Jahre verbesserte Braunrostresistenz. Einer der eingeführten QTL, *QLr.sfr-1BS*, war vorher auf ein Interval von 16 cM auf Chromosom 1BS in ‚Forno' eingegrenzt worden. Weil keines der bereits beschriebenen Braunrostresistenzgene in dieser Region idenitifiziert worden war, erachteten wir den Locus *QLr.sfr-1BS* als neu und benannten ihn *Lr75*.

Um die *Lr75*-Zielregion weiter mit molekularen Markern zu sättigen, verwendeten wir vier verschiedene Strategien: 1) Illumina Survey Sequenzen des Weizenkultivars ‚Chinese Spring' zur Entwicklung von Chromosom 1BS-spezifischen Mikrosatellitenmarker, 2) die flow-sortierten und sequenzierten 1B Chromosomen von ‚Arina' und ‚Forno', 3) den Weizen 90K SNP Array und 4) BAC End Sequenzen (BES) des Chromosoms 1BS von ‚Chinese Spring'. Mittels dieser vier Strategien konnten total 98 Microsatellit- und 146 SNP-basierte Primer entwickelt werden. Aus diesen wurden sechs Mikrosatellit- und 68 SNP-Marker in einer Near Isogenic Line (NIL) Population, bestehend aus 65 BC_3F_2 Linien, kartiert. Die hochauflösende genetische Kartierung der neu entwickelten Mikrosatellit- und SNP-Marker resultierte in der Eingrenzung des *Lr75*-Zielintervals von 16 cM auf 1.2 cM, neu flankiert von den Markern *B2g38480* und *B2g38540a*.

Das entsprechende 3 Mb grosse physische Zielinterval wurde identifiziert durch den Vergleich der Sequenzen der flankierenden Marker mit den Chromosom 1BS-spezifischen BES von BAC Klonen, die mittels Finger Printed Contig (FPC) und Linear Topological Contig (LTC) Assembly von ‚Chinese Spring' erzielt wurden. Die 24 BAC Klone, welche das ganze 3 Mb physische Interval umfassten, wurden komplett sequenziert. Die genetische Kartierung in der NIL Population erfolgte mit

Hilfe von BAC Klonen entwickelter genassoziierter SNP Marker. Aus total 24 dieser neu entwickelten SNP Marker waren 15 polymorph und ihre Kartierung resultierte in der Etablierung eines weiteren verkleinerten 0.96 cM Zielintervals, welches 723 kb in ‚Chinese Spring' entsprach. Zusätzlich wurde ein bisher noch nicht komplettes 403.6 kb physikalisches Interval des *Lr75*-Donors ‚Forno' etabliert. Dies geschah mit Hilfe einer ‚Forno' non-gridded BAC Library, die eine einfache Abdeckung aufweist. Das physikalische Zielinterval in ‚Forno' beinhaltet eine Lücke von unbekannter Grösse, die durch ‚Forno' BAC Library Screening mit neuen Primern geschlossen werden soll.

Durch das manuelle Assembly und der Annotation des physischen *Lr75*-Zielintervals beider Haplotypen wurden 14 putative Kandidatengene in ‚Chinese Spring' und deren 8 in ‚Forno' identifiziert. Der Vergleich der beiden Haplotypen, ‚Chinese Spring' und ‚Forno', enthüllte eine Serie von Insertions-, Deletions- und Duplikationsevents, was auf eine beachtliche Divergenz zwischen den beiden Haplotypen schliessen lässt.

Aus dieser Studie folgern wir, dass 1) Marker-assistierte Kombination von partiellen ‚adult plant' Resistenzgenen eine brauchbare Strategie zur Erhöhung der Breitspektrums-Krankheitsresistenz in Weizen darstellt und 2) die in dieser Studie präsentierten Strategien eine schnelle und effiziente Entwicklung genomspezifischer und genassoziierter SNP Marker zur hochauflösenden genetischen Kartierung sowie für Map-based Cloning ermöglichen.

Chapter 1

General Introduction

1.1 Importance of wheat in agriculture, wheat evolution and domestication

1.2 Diseases of wheat and wheat leaf rust

1.2.1 Leaf rust, *Puccinia triticina* - an obligate biotrophic wheat fungal pathogen

1.2.2 Importance of leaf rust as a wheat disease

1.3 Disease resistance in crop plants

1.4 Genetic mapping and map-based cloning in wheat: Concepts and challenges

1.5 Technical advances to facilitate map-based cloning in wheat

1.5.1 Use of syntenic information between wheat and related model grass species

1.5.2 Advances in wheat genome sequencing

1.5.3 Advancements in molecular marker techniques

1.5.4 Applications of NGS in wheat

1.6 Aim of the thesis

1.1 Importance of wheat in agriculture, its evolution and domestication

Wheat (*Triticum aestivum* L.), being a rich source of proteins, minerals and carbohydrates is one of the most important cereal crops of the world. It serves as a staple food for 40% of the world's population (Peng et al. 2011). The global wheat production amounts to more than 713 million tons, and ranks third after maize (1016 Mt) and rice (745 Mt.). In Europe, wheat was the most produced cereal commodity in 2013 with 225 million tons, followed by maize (117 Mt) and barley (86 Mt). These data are based on statistics from the Food and Agricultural Organization of the United Nations Statistics Division (FAOstat 2014).

Hexaploid bread wheat is the most prominent member of the tribe *Triticeae* which also includes diploid and tetraploid wheat as well as other grasses such as rye (*Secale cereale* L.) and barley (*Hordeum vulgare* L.) (Huang et al. 2002). The origin of the *Triticeae* crops lies in the Fertile Crescent, an arid region spanning the Levante, Mesopotamia and West Iran. During the Neolithic Revolution 10'000 years BP, domestication of wild grasses such as *Tricitum boeticum, Triticum turgidum* spp. *dicoccoides*, led to cultivated grasses such as einkorn (*Triticum monococcum* L.) and emmer wheat (*Triticum dicoccum* L.) (Salamini et al. 2002; Charmet 2011). The selection of advantageous characteristics led to the features of modern bread wheat, which is characterized as the 'domestication syndrome' (Dubcovsky and Dvorak 2007). The 'domestication syndrome' includes the transition from shattering to non-shattering spikes, the acquisition of soft glumes, larger seed size and reduced dormancy, a more erect growth and decreased tiller number.

Hexaploid bread wheat (*Triticum aestivum* L., AABBDD, 2n=6x=42) arose from hybridization of allotetraploid cultivated emmer wheat (*Triticum dicoccum* L, 2n=4x=28, AABB) and the diploid wild goat grass *Aegilops tauschii* L. (2n=2x=DD), approximately 9'000 years ago. The wild form of cultivated tetraploid emmer wheat (*Triticum dicoccum* L, 2n=4x=28, AABB) originated 0.5 million years ago (Huang et al. 2002) from a hybridization event of the wild A genome donor, *Triticum urartu* ($2n = 2x = A^uA^u$) with an unknown B genome donor, a member of the Sitopsis family whose closest modern relative is *Aegilops speltoides* L. (SS).

With the continuing increase of the human population, agricultural production needs an increase of 110% in the next 50 years (Tilman et al. 2011; Alexandratos and Bruinsma 2012). Due to this increase in the world population, the demand for wheat will also grow and a great need lies in increasing the yield potential and actual yield of wheat to ensure food supply. Yield potential refers to the yield which is attained by best cultivars and agronomical practices free from any kind of biotic and abiotic stress. Major constraints reducing actual yield are biotic and abiotic stresses which lead to yield loss. Oerke et al. (1994) estimated an overall yield loss of 12.4% due to all diseases of wheat. Fungicides have been widely used to protect crop plants against diseases and insect pests. Constant use of fungicides is harmful for the environment and has resulted in the emergence of pathogen races that

are resistant to fungicides. Hence, breeding for resistant cultivars provides a more economical and environmental friendly approach to achieve resistance in crop plants.

1.2 Diseases of wheat and wheat leaf rust

Wheat is under constant attack by many pests and pathogens. The most important are the biotrophic foliar fungal pathogens such as rusts (leaf, stripe and stem rust), powdery mildew and smut fungi. In addition, hemibiotrophic fungal pathogens such as Fusarium Head Blight (FHD), tan spot, spot blotch and *Septora* species also attack the wheat crop. Also, soil borne root rots and nematodes cause substantial reduction in wheat yield. Among all these pathogens, the rust fungi are the most important class of wheat pathogens. Rust pathogens consist of more than 7,000 species and belong to the group of *Pucciniales*, Basidiomycota (Duplessis et al. 2011). Wheat rusts are obligate biotrophic pathogens and require a specific host to complete their life cycle. For completing their sexual life cycle, the wheat leaf, stripe and stem rust pathogens require an alternate host. For stripe and stem rusts, the alternate hosts are the species of *Berberis vulgaris* (Jin 2011), whereas for wheat leaf rust the predominant alternate host is *Thalictrum speciosissimum* (Bolton et al. 2008). Although alternative hosts are important for increasing the source of inoculum, the infections at the epidemic scale can occur only in the primary host (Kolmer 2005; Jin et al. 2010). In case of wheat leaf rust, the causal agent is *Puccinia triticina*, whereas the causal agents of stem and stripe rusts are *Puccinina graminis* and *Puccinia striiformis*, respectively. Infectious rust spores can travel long distances and spread the disease through wind dissemination and under favourable conditions can cause severe epidemics on a global scale (Brown and Hovmøller 2002; Bolton et al. 2008). Wheat leaf rust is the most widely distributed and most frequently occurring rust of wheat that causes substantial yield losses annually which can range from 14% to 40% (Bolton et al. 2008; Huerta-Espino et al. 2011; Kolmer 2013). Under epidemic conditions, yield losses in stripe rust were reported to be in the range of 10-70% (Chen 2005). The emergence of stem rust race Ug99 increased the awareness of stem rust as a disease (Singh et al. 2011).

1.3 Leaf rust (*Puccinia triticina*) - an obligate biotrophic wheat fungal pathogen

Being an obligate biotrophic pathogen, *P. triticina* requires a living host to grow and complete its life cycle. The rust fungi while invading the host form highly specialized infection structures that are involved in spore attachment, host recognition, penetration, proliferation and nutrition (Mendgen and Hahn 2002). Haustorium is the most complex infection structure developed by rust fungi while invading the host cell. It serves as a feeding structure to extract nutrients from the host cell towards the pathogen and also leads to the suppression of the plant's triggered defense responses. The infection process starts when air-borne infectious rust spores land on the host plant and start to germinate. The

germination of the infectious spores leads to the formation of a primary germ tube which is directed towards the stomata of the leaf by thigmotrophic growth and results in the formation of an appressorium over the stomatal aperture. After formation of an appressorium, an infection peg grows into the substomatal cavity. From the infection peg, infection hyphae start to develop which upon contact with mesophyll cells form haustorial mother cells which further develop haustoria. Haustoria then invaginate the mesophyll plasma membrane by penetrating through the cell wall (Webb and Fellers 2006) (Fig. 1.1). After the successful establishment of haustorial structures and uptake of nutrients, orange-coloured infectious pustules (urediniospores) appear on the leaf surface usually 7-10 days after infection. A new infection cycle starts with water or wind dispersed urediniospores.

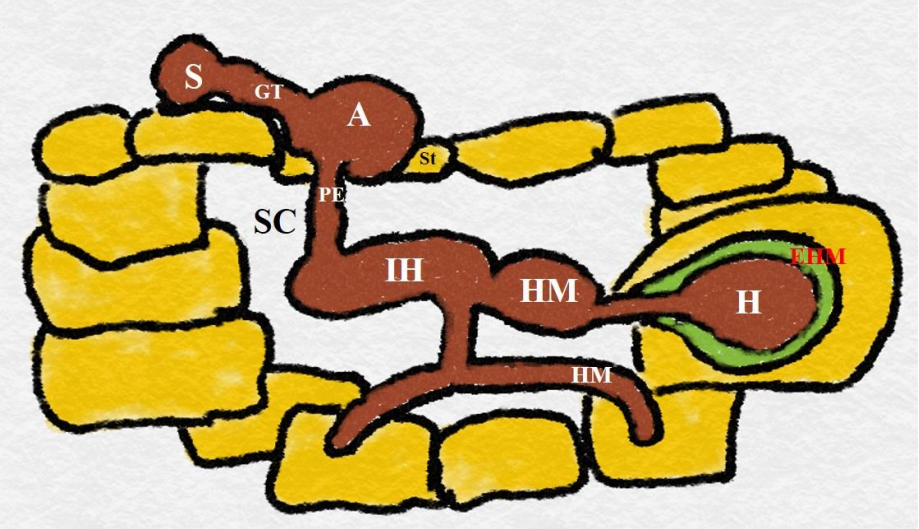


Fig. 1.1 Infection structures of the obligate biotrophic fungus, *Puccinia triticina* (leaf rust). A: appressorium, S: spore, GT: germination tube, PE: penetration peg, IH: infection hyphae, HM: haustorial mother cell, H: haustorium, EHM: extrahaustorial matrix, SC: substomatal cavity, St: stomata.

Fig. 1.2 Wheat leaf rust on the wheat flag leaf is visible as medium to large sized, orange coloured uredinia

1.3.1 Economic importance of wheat leaf rust

Leaf rust is the most widely distributed and most frequently occurring rust of wheat that causes substantial yield losses annually (Bolton et al. 2008; Huerta-Espino et al. 2011; Kolmer 2013). Leaf rust mainly attacks the leaf blades but in severe cases the infection can spread to leaf sheath and glumes. Early infection of 60-70% on the flag leaves during spike emergence can cause 30% or more yield loss. On the other hand, the yield loss is only 7% if the same level of infection occurs at soft dough stage (Huerta-Espino et al. 2011). Yield losses due to leaf rust are characterized by reduced kernel weight and lower number of kernels per head (Bolton et al. 2008; Huerta-Espino et al. 2011).

1.4 Disease resistance in crop plants

Disease resistance in plants has been categorized into several categories based on the nature of genetic control, durability, specificity, stage of infection and level of infection. When resistance is controlled by either single or multiple genes towards several, different pathogen races, it is called

monogenic or polygenic resistance, respectively. Based on the spectrum of specificity, resistance can be either race-specific or race non-specific. When disease resistance is conferred to only a particular or some races of a pathogen, it is known as race-specific resistance. On the other hand, race non-specific resistance genes confer resistance against all races of a pathogen. The third classification of disease resistance based on durability, categorizes resistance genes as durable or non-durable. Resistance provided at all stages from seedling to adult plant is known as seedling resistance whereas the resistance at the adult plant stage only is called adult plant resistance (APR). The majority of the rust resistance genes are seedling resistance genes because of their effectiveness at all the growth stages of the plant. Based on the infection level, resistance can be complete which leads to the death of the cell under attack; or partial where the pathogen can develop to reduced level, reducing the pressure on the pathogen to evolve.

Race-specific resistance is typically under monogenic control and is often overcome by emerging virulent pathogen races. Hence, race-specific resistance is mostly non-durable in the field. Race-specific resistance involves an interaction of a specific effector or avirulence gene (*Avr* gene) of the pathogen with a specific resistance (*R*) gene in the plant. Fungal effectors (Avr proteins) are the proteins which either suppress or induce the plant immune system. Fungal effectors which are recognized by host resistance (*R*) proteins are known as avirulence factors (*Avr*) and trigger the resistance response in the plant. The presence of the avirulence gene in the pathogen and the *R* gene in the plant leads to an incompatible interaction followed by a hypersensitive response (HR) and subsequent cell death. Flor (1955) was the first to describe this "gene-for-gene" interaction in flax. A number of *R* genes have been cloned and most of them belong to NBS-LRR family (nucleotide binding site and leucine rich repeat domains) (Wang et al. 2011). The powdery mildew resistance gene *Pm3* is an example of a cloned race-specific resistance gene against the powdery mildew pathogen *Blumeria graminis* f. sp. *tritici* (Yahiaoui et al. 2004). Also, several genes which provide resistance against wheat leaf rust (*Puccinia triticina*) and belong to the NBS-LRR class of proteins have been cloned; *Lr21, Lr10* and *Lr1* (Huang et al. 2003; Feuillet et al. 2003; Cloutier et al. 2007). This kind of resistance can be overcome by mutations in the fungal *Avr* gene disabling its recognition by the *R* gene of the plant. For example, the wheat stem rust resistance gene *Sr24*, which was introgressed from *Agropyron* spp., was overcome by virulent pathogen races. Similarly, the resistance provided by the stem rust resistance gene *Sr31*, translocated from rye (*Secale cereale* L.), was overcome by the African stem rust race Ug99 (originating in Uganda 1999). With the emergence of the Ug99 lineage, 90% of the world's wheat cultivars became susceptible to stem rust, threatening 20% of the wheat production in Asia, the Middle East and Central and North Africa (Dixon et al. 2009; Singh et al. 2011). However, a few monogenic, resistance genes have also been identified and cloned in wheat that are race non-specific and confer broad-spectrum resistance. For example, the two stem rust (*Puccinia graminis* f. sp *tritici*) resistance genes *Sr33* and *Sr35* confer broad-spectrum resistance to wheat stem rust (Periyannan et al. 2013; Saintenac et al. 2013b).

Race non-specific resistance, on the other hand, is effective against all races of a given pathogen species. Such type of resistance is mainly effective at the adult plant stage and is also durable in nature because of the low growth rate of the pathogen which reduces the selective pressure (Lindhout 2002). Selection of wheat cultivars with durable disease resistance has been the major objective of breeding programmes to ensure global food security. Such wheat cultivars have been incorporated into the wheat germplasm by the International Maize and Wheat Improvement Centre CIMMYT. Race non-specific resistance is often quantitative in nature, and therefore attributed to a set of minor genes or quantitative trait loci (QTLs). The resistance is mainly partial without HR. Quantitative resistance in a plant may be expressed as delayed and/or reduced growth of the pathogen compared to a susceptible plant, that is why such resistance genes against leaf rust are also called slow-rusting genes (Ribeiro Do Vale et al. 2001). The genetic complexity of the quantitative resistance complicates the map-based isolation of disease resistance genes in crops. Nevertheless, three rust resistance QTLs in wheat have been cloned: the leaf rust resistance gene *Lr34* (Krattinger et al. 2009), *Lr67/Yr46/Sr55/Pm46* (Moore et al. 2015) and the stripe rust resistance gene *Yr36* (Fu et al. 2009). Apart from these, other sources of durable APR leaf rust resistance genes have been genetically identified in wheat: *Lr46/Yr29/Pm39* (Singh et al. 1998; Rosewarne et al. 2006; Lillemo et al. 2008) and *Lr68* (Herrera-Foessel et al. 2012) as well as the stem rust resistance gene *Sr2* (Singh et al. 2008). APR genes should not always be correlated with race non-specificity as there are some APR genes which are race-specific in nature. Examples are the leaf rust resistance genes *Lr12, Lr13, Lr22a, Lr22b* and *Lr37* (Mcintosh et al. 1995). None of these genes has been cloned yet.

1.5 Genetic mapping and map-based cloning in wheat: Concepts and challenges

Genetic mapping and map-based cloning, also known as positional cloning is a strategy to isolate the gene of interest when no previous knowledge of the gene product is available. The essential step for map-based cloning is the establishment of a mapping population which is derived from crossing two cultivars differing for the trait of interest. Afterwards, a precise genetic map is constructed by the combination of genetic marker data and phenotypic evaluation, therefore indicating the position of the gene of interest. Then the closest flanking markers are used to define the physical region spanning the gene of interest by screening BAC libraries and chromosome walking (Krattinger et al. 2007). Further sequencing of the complete physical region leads to the identification of candidate genes. The establishment of a physical target region and identification of candidate genes finally lead to the last step in map-based cloning, which is the validation of candidate genes. Several techniques are used to validate candidate genes, such as mutant analysis, allelic diversity studies, expression studies, stable transformation assays and virus-induced gene silencing (VIGS). For instance, Krattinger et al. (2009) cloned the leaf rust gene *Lr34* present on wheat chromosome 7DS by generating loss-of-function mutations. Similarly, the stripe rust resistance gene *Yr36* was cloned by Fu

et al. (2009) with the help of EMS-induced mutants. Advancements in Targeting Induced Local Lesions in Genomes (TILLING) complements classical mutagenesis. It involves the combination of chemical mutagenesis with high throughput screening of mutant populations (Uauy et al. 2009).

Map-based cloning in wheat, however, is a challenging task due to several factors. One of them is the enormous genome size of wheat, which is approximately 17 Gb. A second challenge is the repetitive nature of the wheat genome which consists mainly of retrotransposons and DNA transposons (Wicker et al. 2003). Due to the presence of repetitive elements, marker development and chromosome walking steps become difficult in map-based cloning. A third difficulty is the presence of three homeologous sub-genomes that contain three very similar copies of the same gene, therefore requiring extra work to identify the correct sub-genome. Despite of these difficulties, a handful of *R* genes have been identified and isolated in wheat. Most of the cloned disease resistance genes in wheat belong to the CC-NBS-LRR group. However, a more challenging task is the map-based cloning of durable resistance genes in wheat because of their partial effect and polygenic nature. For dissection of such complex traits, QTL studies are required. Nonetheless, efforts have been undertaken to identify and clone durable, quantitative disease resistance genes in wheat. The most successful examples are the isolation of leaf rust and stripe rust resistance QTLs, *Yr36, Lr34* and *Lr67* in wheat (Fu et al. 2009; Krattinger et al. 2009; Moore et al. 2015). The advancements mentioned below have facilitated genetic mapping in crops with complex genomes such as wheat.

1.6 Technical advances to facilitate map-based cloning in wheat

1.6.1 Use of syntenic information between wheat and related model grass species

Among cereals, hexaploid bread wheat has the largest genome size with 17’000 megabase pairs (Mbp) which is 3-fold and 35-fold larger than that of barley (*Hordeum vulgare*, 5500 Mbp) and rice (*Oryza sativa* L., 420 Mbp), respectively. Polyploidy provides an enormous versatility in terms of growth conditions, resistance to diseases and the potential to process the harvested seeds to various products. However, the domestication of wheat resulted in a diversity bottleneck (Dubcovsky and Dvorak 2007). The similarity at the genomic level among the modern wheat genomes amounts to 99.9 % (Chao et al. 2009; Trick et al. 2012). An important source of new diversity in wheat is the activity of the highly abundant transposable elements (TEs) which compose 80% of the genome. Many gene deletions and mutations can be attributed to the action of TEs. Together with frequent gene duplication caused by unequal crossing-over or chromosome breakage during replication, TEs account for a highly dynamic wheat genome (The International Wheat Genome Sequencing Consortium 2014). However, gene order and synteny is still conserved among the three sub-genomes A, B and D, therefore revealing high sequence similarity with only a small amount of sub-genome specific genes (Dubcovsky and Dvorak 2007).

The grass family consists of a diversity of species including wheat (*Triticum aestivum* L., 2n=6x=42, 17 Gb), barley (*Hordeum vulgare* L., 2n=2x=14, 5.1 Gb), maize (*Zea mays* L., 2n=2x=20, 2.5 Gb), sorghum (*Sorghum bicolor* L., 2n=2x=20, 730 Mb), rice (*Oryza sativa* L., 2n=2x=24, 400 Mb), and *Brachypodium distachyon* L. (2n=2x=10, 273 Mb) which show variation in terms of genome size, chromosome number and ploidy level (Devos 2005; Eckardt 2008). The gene order in these species is partially conserved due to their common divergence time of about 25-77 MYA (Eckardt 2008; Wicker et al. 2011). The presence of syntenic sets of genes between wheat and other grass genomes is very useful for genetic mapping studies in wheat. Synteny refers to the conservation of gene order between different chromosomal regions among different genomes. The syntenic information from the complete sequenced genomes of rice, sorghum, maize, *B. distachyon* and barley, can be used to deduce the hypothetical gene order and gene content in wheat (International Rice Genome Sequencing Project 2005; Paterson et al. 2009; Schnable et al. 2009; Vogel et al. 2010; Mayer et al. 2012). By comparing the syntenic gene regions, a reference "genome zipper" has been created for wheat and barley chromosomes (Mayer et al. 2011; Wicker et al. 2011; Hernandez et al. 2012). In addition, syntenic information between wheat and sequenced grass genomes is also useful for estimating the order of genetic markers and estimation of the number of putative candidate genes in the target region. However, breaks in the syntenic gene order between wheat and different grass genomes in terms of small inversions, deletions and duplications have also been observed (Akhunov et al. 2003; Raats et al. 2013). Studies have also shown that the presence of non-syntenic genes resulted in the breakdown of synteny in wheat (Choulet et al. 2010; Raats et al. 2013). However, with the advancements in genome sequencing of diploid wheat progenitors and hexaploid wheat (Ling et al. 2013; Jia et al. 2013; The International Wheat Genome Sequencing Consortium 2014), it might be possible that model grass genomes might not be needed anymore for positional cloning in wheat in the near future.

1.6.2 Advances in wheat genome sequencing

Before the advent of next generation sequencing (NGS) technologies, the main focus of wheat genome analysis was on the study of coding sequences which cover less than 2% of the wheat genome. Efforts by scientific community resulted in the generation of 40,000 unigenes, more than 1 million expressed sequence tags (ESTs), out of which nearly 7,000 ESTs were assigned to different wheat chromosomes (http://www.ncbi.nlm.nih.gov/genbank/dbest/dbest_summary/), and 17,000 full-length complementary DNA (cDNA) sequences (Mochida et al. 2009). All these resources were extensively used in marker development for mapping as well as gene expression studies (Bernardo et al. 2009; Chelaifa et al. 2013).

With the emergence of next generation sequencing (NGS) technologies, efforts have been undertaken by the International Wheat Genome Sequencing Consortium (IWGSC) and others to

achieve the complete sequence of the wheat genome. Brenchley et al. (2012) used a whole genome shotgun approach (454 pyrosequencing) to sequence the genome of the wheat cultivar 'Chinese Spring'. Although this genome was highly fragmented, this approach is still highly useful in genomics studies. They identified 94,000- 96,000 genes which were assigned to the three A, B and D sub-genomes of hexaploid wheat, also exploiting the ancestral genome sequences of *Triticum monococcum, Aegilops speltoides* and *Aegilops tauschii*. However, whole genome shotgun sequencing has been more successful in sequencing and assembling the small and less complex genomes of soybean (*Glycine max,* 1.1 Gb) and potato (*Solanum tuberosum,* 844 Mb) (Schmutz et al. 2010; Xu et al. 2011). Apart from this, other approaches have been used to access the hexaploid wheat genome. One of them is to singly access each of the three sub-genomes by sequencing the diploid relatives *T. urartu, Ae. speltoides* and *Ae. tauschii.* Therefore, BAC libraries of these three species were generated (Akhunov et al. 2005) and more recently, the draft genome sequences of *T. urartu* and *Ae. tauschii* were completed (Ling et al. 2013; Jia et al. 2013).

To further reduce the complexity of the hexaploid wheat genome, the individual chromosomes and chromosome arms of the wheat cv. 'Chinese Spring' were isolated using flow cytometry sorting (Šafář et al. 2004; Safár et al. 2010; Doležel et al. 2012; Vrána et al. 2015). BAC libraries designed from the flow-sorted chromosomes allowed the construction of physical maps and sequencing data for each individual chromosome arm (Paux et al. 2008; Doležel et al. 2012; Lucas et al. 2013; Raats et al. 2013) These chromosome specific BAC libraries are now available at CNRGV, France. Recently, a chromosome-based draft sequence of the hexaploid bread wheat was completed having 0.41 to 0.89 times fold coverage range (The International Wheat Genome Sequencing Consortium 2014). A total of 124,201 gene loci were annotated which were evenly distributed throughout the homeologous chromosomes and the three sub-genomes of wheat. The reference genome sequence of individual chromosomes can be accessed at the IWGSC sequence Repository webpage (http://wheat-urgi.versailles.inra.fr/Seq-Repository). All those resources provide a great deal of advantages for marker development and for the identification of chromosome-specific sequences using BLAST searches. Furthermore, they simplify the identification of genes based on syntenic information from the sequences of model grass genomes.

1.6.3 Advancements in molecular marker techniques

Molecular markers present one of the most important resources in genetic mapping and map-based cloning studies. During the past two decades, molecular markers have been efficiently used in various wheat breeding programmes. Several national and international mapping projects have been initiated such as the International Triticeae Mapping Initiative (ITMI) to generate a large set of molecular markers (http://wheat.pw.usda.gov/ITMI/).

In the past years, the development of molecular markers was very laborious and time consuming due to the lack of sequence information. Classical markers used during the past decades were restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP) and Diversity Arrays Technology (DaRT). All these markers did not require any prior sequence information and were therefore widely used in genetic mapping and genetic diversity studies (Jonah et al. 2011; Paux et al. 2012). Unfortunately, due to only low through-put detecting techniques, they were only partially amenable to molecular breeding applications.

With the introduction of Sanger sequencing and various developments in NGS technologies, the "second generation" of molecular markers which rely on sequence information came into play which mainly includes simple sequence repeats (SSR), also known as microsatellite markers and EST-based markers. SSR markers are short tandem repeat sequences and are designed from repetitive regions of the genome. ESTs are low copy, expressed sequences in the genome. Numerous efforts have been undertaken during the last ten years to produce wheat ESTs. Over 1 million ESTs are now present in the EST database, dbEST (http://www.ncbi.nlm.nih.gov/dbEST/index.html). Both classes of markers are reviewed in Paux et al. (2012)

With the increasing availability of sequencing resources of the wheat genome, single nucleotide polymorphisms (SNP) and insertion site based polymorphisms (ISBP), created by the movement of transposable elements (TEs) in the genome (Paux et al. 2010; Paux et al. 2012), are becoming the markers of choice. Both SNPs and ISBPs are not only highly abundant throughout the genome, but also tractable to high throughput genotyping platforms such as high resolution melting (HRM), fluorescent capillary electrophoresis and Kompetitive Allele Sequencing PCR (KASP).

1.6.4 Applications of NGS in wheat

The wheat genomic approaches described above have facilitated the identification and isolation of genes of interest. However, SNP discovery, high resolution mapping and map-based cloning is still much easier and more straightforward in crops for which a complete reference genome sequence is available. This is for example the case for rice and maize. For instance, RNA-sequencing discovered nearly 5,000 valid SNPs in approximately 2,500 genes in rice and maize (Barbazuk et al. 2007; Zhang et al. 2010). In rice, SNP based linkage maps were constructed by Xie et al. (2010). In maize, candidate genes were identified through SNP discovery (Buckler et al. 2009). Huang et al. (2009) identified 1,493,461 SNPs by resequencing 50 recombinant inbred lines (RILs) of rice through Illumina technology. In wheat, Wang et al. (2014) generated a 90,000 SNP array which was used to characterize the genetic diversity in tetra- and hexaploid wheat. From this 90K SNP array, 46,977 SNPs were genetically mapped using eight different mapping populations. SNPs were also used in genome wide association studies (GWAS) applied on a worldwide collection of 1,000 hexaploid

spring wheat cultivars to identify the sources of resistance to wheat stripe rust (Maccaferri et al. 2015). They identified three new QTLs for stripe rust resistance which mapped away from the already identified and described stripe rust resistance genes. The combination of NGS technologies and the development of SNP-based markers facilitated the identification of polymorphisms between different cultivars and therefore, led to the construction of high density genetic linkage maps. Furthermore, the availability of BAC-based reference sequences and physical maps of the flow-sorted individual chromosome or chromosome arms facilitates the identification of BAC clones harbouring the gene of interest and help in establishing the physical region of interest. For example, Mago et al. (2014) exploited the physical map information of wheat chromosome 3B and constructed a chromosome-specific BAC library using the technique of flow-cytometry sorting to define the physical region for stem rust resistance gene, *Sr2* in wheat. Using these techniques, they identified a physical region of 567 kb in the wheat cv. 'Chinese Spring' and 867 kb in the resistant cultivar 'Hope'. Now days with NGS techniques, to validate the candidate genes of interest, exome capture platform is gaining popularity which involves the sequencing of the coding region of the genome, also known as exome in the mutagenized populations. Uauy et al. (2009) used this technique for capturing DNA from the mutants of tetraploid wheat. Henry et al. (2014) in their study identified mutations by combined use of exome capture and NGS, by using mutant populations of wheat and rice.

1.7 Aim of the thesis

The aim of the study was the identification, characterization and high-resolution genetic and physical mapping of the novel leaf rust resistance gene *Lr75* present on chromosome 1BS in the Swiss winter wheat cv. 'Forno'. The first chapter focuses on the identification and characterization of *Lr75* as a partial adult plant leaf rust resistance gene which was initially described as a QTL in an 'Arina x Forno' recombinant inbred line (RIL) population. The second chapter mainly highlights different marker approaches used during the course of this study to map *Lr75* at high resolution. The focus of the third chapter is on the identification of the physical target region spanning *Lr75* using the 1BS physical information of wheat cv. 'Chinese Spring' and flow-sorted survey sequence information.

Chapter 2

Characterization of *Lr75* – a partial, broad-spectrum leaf rust resistance gene in wheat

Jyoti Singla, Linda Lüthi, Thomas Wicker, Urmil Bansal, Simon G. Krattinger and Beat Keller

Submitted in TAG (Theoretical and Applied Genetics)

2.1 Summary

Leaf rust caused by the fungal pathogen *Puccinia triticina* is a damaging disease of wheat (*Triticum aestivum* L.). The combination of several, additively-acting partial disease resistance genes has been proposed as a suitable strategy to breed wheat cultivars with high levels of durable field resistance. The Swiss winter wheat cultivar 'Forno' continues to show near-immunity to leaf rust since its release in the 1980s. This resistance is conferred by the presence of at least six quantitative trait loci (QTL), one of which is associated with the morphological trait leaf tip necrosis (LTN). Here, we used a marker-informed strategy to introgress two 'Forno' QTLs into the leaf rust-susceptible Swiss winter wheat cultivar 'Arina'. The resulting backcross line 'Arina*LrFor*' showed markedly increased leaf rust resistance in multiple locations over several years. One of the introgressed QTL, *QLr.sfr-1BS*, is located on chromosome 1BS. We developed chromosome 1B-specific microsatellite markers by exploiting the Illumina survey sequences of wheat cv. 'Chinese Spring' and mapped *QLr.sfr-1BS* to a 4.3 cM interval flanked by the SSR markers *gwm604* and *swm271*. *QLr.sfr-1BS* does not share a genetic location with any of the described leaf rust resistance genes present on chromosome 1B. Therefore, *QLr.sfr-1BS* is novel and was designated as *Lr75*. We conclude that marker-assisted combination of partial resistance genes is a feasible strategy to increase broad-spectrum leaf rust resistance. The identification of *Lr75* adds a novel and highly useful gene to the small set of known partial, adult plant leaf rust resistance genes.

2.2 Introduction

Hexaploid bread wheat (*Triticum aestivum* L.) is one of the three most important cereal crops with an annual global production of 713 million tonnes (FAOSTAT 2013, http://faostat3.fao.org). Wheat is attacked by many pathogens of which the fungal rust diseases are the most widespread and devastating. There are three species of wheat rust: leaf or brown rust (*Puccinia triticina*), stripe or yellow rust (*P. striiformis* f. sp. *tritici*) and stem or black rust (*P. graminis* f. sp. *tritici*). Leaf rust is the most common and most widespread rust disease (Bolton et al. 2008; Kolmer 2013). Yield losses caused by leaf rust are characterized by reduced kernel weight and a lower number of kernels per spike (Bolton et al. 2008; Huerta-Espino et al. 2011).

The release of crop varieties with high levels of durable disease resistance represents the most sustainable strategy to reduce production losses caused by fungal diseases. To date, 74 leaf rust resistance (*Lr*) genes have been described (McIntosh et al. 2014). Most *Lr* genes confer race-specific seedling resistance. Rapid pathogen adaptation often results in the emergence of new virulent leaf rust races and consequently a breakdown of race-specific *Lr* resistance (Huerta-Espino et al. 2011; Mcintosh et al. 2014). Hence, there is a need to identify *Lr* genes that show a more durable resistance in the field. A particular type of durable disease resistance is referred to as slow-rusting resistance. It is characterized by a partial resistance phenotype that is often only effective at the adult plant stage but not in seedlings (Caldwell 1968). This type of resistance is therefore also referred to as adult plant resistance (APR). Because of their partial nature it is challenging to combine several APR genes in a single genotype through classical breeding. Marker-assisted selection can serve as a suitable approach to track the presence of APR loci in breeding programs. It has been described that the combination of two or more APR loci with additive effects can result in near immune resistance levels (Singh et al. 2000, Lillemo et al. 2011).

Wheat breeders at the International Wheat and Maize Improvement Center (CIMMYT) have exploited the strategy of combining slow-rusting genes in wheat disease resistance breeding. The CIMMYT wheat breeding material has therefore been the most important source for the discovery of slow-rusting genes. Singh et al. (2005) have shown that durable adult plant resistance in CIMMYT bread wheats is due to the presence of at least 10-12 slow-rusting genes present in different combinations. Some of these genes are mapped or cloned at high resolution such as *Lr34/Yr18/Sr57/Pm38* and *Lr68* (Lagudah et al. 2006; Krattinger et al. 2009; Herrera-Foessel et al. 2012). On the other hand, slow-rusting resistance has only been poorly studied in other wheat material such as the Central European winter wheat genepool.

Due to its large (17 Gb) and repetitive genome (80%), genetic mapping in hexaploid wheat is a challenging task (Wanjugi et al. 2009; Lucas et al. 2012). In addition, the development of molecular markers is a tedious and time consuming process because of the low level of polymorphism between elite wheat cultivars (Paux et al. 2012). Despite these difficulties, molecular markers have been widely

used in the identification and mapping of rust resistance genes. With the advent of new technologies in wheat genomics, the development of polymorphic markers has become easier and faster. A key breakthrough in reducing the complexity of the wheat genome was the sorting of individual chromosomes by flow-cytometry which facilitates the construction of chromosome-specific BAC libraries and the development of physical maps (Janda et al. 2004; Safár et al. 2010). In addition, the use of completely sequenced genomes of closely related diploid grasses such as *Brachypodium*, rice and barley as model organisms also facilitates the characterization of the wheat genome by comparative approaches (International Rice Genome Sequencing Project 2005; Vogel et al. 2010; Mayer et al. 2012).

The Swiss winter bread wheat cultivar 'Forno' (pedigree: 'NR72837 x Kormoran') was released in Switzerland in 1986. 'Forno' shows near immune leaf rust resistance in the field against all leaf rust isolates tested so far. At least six quantitative resistance loci contribute to this remarkable leaf rust resistance of 'Forno' (Schnurbusch et al. 2004). Among them is the well-known multi-pathogen resistance gene *Lr34* located on chromosome 7D. *Lr34* encodes for an ATP-binding cassette transporter (Krattinger et al. 2009) and it explained 33-43% of the phenotypic variation in the QTL study of Schnurbusch et al. (2004). *Lr34* is associated with leaf tip necrosis (LTN), a senescence-like process. LTN is often seen as an unwanted trait in Western European wheat cultivars and *Lr34* has therefore only rarely been used in the Western European wheat breeding programs (Kolmer et al. 2008). A second major QTL for leaf rust resistance, *QLr.sfr-1BS* was identified on chromosome arm 1BS. *QLr.sfr-1BS* explained 28-32% of the phenotypic variability and was not associated with LTN. This locus interacted with a minor QTL (*QLr.sfr-7BL*) on chromosome arm 7BL. In a different mapping population derived from 'Forno' crossed with the spelt wheat cv. 'Oberkulmer', Messmer et al. (2000) identified six QTLs for durable leaf rust resistance in cultivar 'Forno'. Three of these QTLs were not detected by Schnurbusch et al. (2004). The strongest QTL detected in the 'Forno' x 'Oberkulmer' population was on chromosome 7BL with a phenotypic variance of 35.8%. This QTL fell into the same genetic interval as a minor QTL identified by Schnurbusch et al. (2004). Also, a QTL on chromosome 1BS was identified explaining a phenotypic variation of 10.6% across four environments. This is most likely be the same QTL as *QLr.sfr-1BS* because it is identified in the same genetic interval as *QLr.sfr-1BS* in 'Arina' x 'Forno' RIL population.

Here, we describe that the marker-assisted introgression of the two leaf rust resistance QTLs *QLr.sfr-1BS* and *QLr.sfr-7BL* from 'Forno' resulted in partial leaf rust resistance in the popular but leaf rust-susceptible Swiss winter wheat cultivar 'Arina'. *QLr.sfr-1BS* was further fine-mapped and was found to be a yet uncharacterized, partial leaf rust APR gene that was designated as *Lr75*. *Lr75* mapped to a 4.3 cM interval flanked by markers *gwm604* and *swm271*.

2.3 Materials and Methods

2.3.1 Plant material

Two different mapping populations were used in this study. The first population consisted of 240 recombinant inbred lines (RIL) generated from a cross of two Swiss winter wheat cultivars, 'Arina' and 'Forno' (Schnurbusch et al. 2004). A second population was developed from a cross of 'Arina' and back-cross line 'Arina*LrFor*'. A near-isogenic back-cross line, 'Arina*Lr34*' (BC_3F_7) with *Lr34* in the genetic background of susceptible cv. 'Arina' was also included in the field trials for comparison. The recombinants of the 'Arina' x 'Arina*LrFor*' population were phenotyped as F_3, F_4 and F_5 rows in the field in Reckenholz, Switzerland during 2013, 2014 and 2015.

2.3.2 Characterization of leaf rust resistance in the field

For field trials in Switzerland, the parental lines, 'Arina', 'Forno, 'Arina*LrFor*', 'Arina*Lr34*' and the recombinants were sown in randomized 5-row plots of 130 cm in two replications with 40 seeds per row. Every 20th plot consisted of the parental lines. The first and the last row in each plot were spreader rows containing a 1:1:1 mixture of highly susceptible wheat lines 'Morocco', 'Bernina' and 'Arina' to facilitate high and uniform pathogen density in the field. All field trials were inoculated with a mixture of 16 Swiss leaf rust isolates as described in Messmer et al. (2000). Leaf rust severity on the flag leaves of the parental lines and the population were recorded when susceptible cv. 'Arina' displayed a leaf rust infection of 70%.

'Arina', 'Forno', 'Arina*LrFor*' and 'Arin*Lr34*' were also tested in the fields in Cobbitty, Australia in 2014. The lines were planted at the Karalee site of the Plant Breeding Institute of the University of Sydney. Leaf rust-susceptible genotype 'Sonora' was used in spreader rows. Urediniospores of *Puccinia triticina* (Pt) pathotype 76-1,3,5,7,9,10,12 +*Lr37* (Plant Breeding Institute Culture Number =621) and 10-1,3,9,10,11, 12 (Plant Breeding Institute Culture Number =592) were used for infections. The virulence/avirulence formulae of *Pt* pathotypes which were used to test the parental lines is provided in Appendix A. Disease severity to leaf rust on the flag leaves of adult plants in the field were recorded when 'Arina' showed an infection of 60%.

2.3.3 Characterization of leaf rust resistance in the greenhouse

The parents 'Arina', 'Forno' and 'Arina*LrFor*' along with 'Arina*Lr34*' were characterized at seedling stage in the greenhouse with the seven Swiss isolates 91047, 96002, 95219, 93012, 96209, 95001 and 90035 which were part of the isolate mixture used for field infections. Virulence/avirulence formulae of these seven isolates is given in Appendix B. For each line, 20 to 25 seeds were sown in two replicates in soil (Rasenerde [20% org. matter, pH ($CaCl_2$) 6.5, 1.4g/L salt content (KCl), filler (DIN EN 12580)], ökohum GmbH, Herrenhof, Switzerland) in pots with a diameter of 13 cm. After

treatment with growth inhibitor (Cycocel® Extra (4ml/L), Omya AG, AGRO, Oftringen, Switzerland) and fertilizer (Wuxal® Profi (2-3ml/L), Maag Garden, Syngenta, Düsseldorf, Germany), they were grown for 10 days under diurnal conditions (16h light/20°C, 8h dark/16°C, 70% humidity). At the two-to-three-leaf stage (approx. after 10 days) the plants were inoculated with a suspension of urediniospores and oil (3M™ Fluorinert™ FC-43, 3M Electronics, Zwijndrecht, Belgium). After inoculation, the plants were allowed to air dry for 30 minutes before they were placed in darkness for 24h at 16°C with 95% humidity. Afterwards, the plants were transferred to growth chambers providing 16h light/20°C, 8h dark/16°C, 70% humidity. The disease was assessed 10 days after infection (dai) using the 0-4 infection type scoring system described by Roelfs et al. (1992).

2.3.4 Measurement of pustule density of urediniospores

Pustule density of the urediniospores was measured on field-infected flag leaves of 'Arina', 'Forno', 'Arina*LrFor*' and 'Arina*Lr34*' when the plants were between the Zadoks growth stage 55-69 i.e. from half ear emergence to complete anthesis (Zadoks et al. 1974). For each line, leaves from two different plots were sampled and nine flag leaves per plot were randomly selected. For each leaf, a surface of 4 centimetres in the middle of the leaf was marked. The marked area was photographed at 3 different time points (84 days, 87 days and 93 days after infecting the spreader rows) with a Nikon camera using a macro objective lens. Subsequently, the pictures were analyzed with the picture analysis program ImageJ 1.48 (Abràmoff et al. 2004). The area of interest (4 cm x width of the leaf) was calculated using the polygon selection tool; the pustules were counted using the oval selection tool. Not fully visible pustules at the edge of the leaf or the marked area were ignored. The average pustule density (number of pustules per area) was calculated for each of the four lines at different time points. Verification of the differences in pustule density between the lines was done by the *t*-test using JMP 11 software (SAS Institute Inc., Cary, NC, 1989-2007).

2.3.5 Marker analysis and genetic linkage mapping

The parents and BC_3F_2 recombinants were grown in the greenhouse and leaf tissue was harvested from 8-10 days old seedlings. DNA was extracted with a CTAB (cetyl trimethyl ammonium bromide) protocol as described by Stein et al. (2001). The quantity and concentration of DNA was measured using a NanoDrop spectrophotometer (Witec AG, Lucerne, Switzerland). The final concentration was standardized to 650 ng/µL. Dilutions of 13 ng/µL and 65 ng/µL were used in PCR reaction for 6% LiCOR gel (LiCOR DNA Sequencer 4200) and agarose gel electrophoresis, respectively. The simple sequence repeat identification tool, SSRIT (http://archive.gramene.org/db/markers/ssrtool) was used to identify the repeat motifs and SSR primers were designed using the software program Primer3 (v. 0.4.0). The PCR products of SSR

primers were resolved on 6% LiCOR gel. SSR markers were named as ‘swm’ (swiss wheat microsatellites). Primer sequences along with their repeat motifs are given in Appendix C. The genetic linkage map was constructed on a subset of F_2-derived F_3 lines (lines with missing phenotypic data were excluded) by calculating the recombination frequency between the markers. MapChart 2.3 (Voorrips 2002) was used to draw the linkage map.

2.3.6 Deletion bin mapping

Chromosome 1B specificity of the SSR markers was confirmed by the absence of amplification in nulli-tetrasomic lines of cultivar ‘Chinese Spring’. Further, to determine the bin localization of all SSR markers, a set of 11 deletion lines for chromosome 1B was used along with two ditelosomic lines of wheat cv. ‘Chinese Spring’. The fraction length (FL) value of each deletion line depicts the length of the remaining chromosome arm from the centromere after deletion relative to the length of the complete arm (Endo and Gill 1996). All the cytogenetic stocks were kindly provided by J. Raupp, Wheat Genetic Resource Centre, Department of Plant Pathology, Kansas State University, USA.

2.4 Results

2.4.1 Selection of 'Forno' leaf rust QTLs for backcrossing

Schnurbusch et al. (2004) identified 6 QTLs for leaf rust resistance in 'Forno' in a population of 240 'Arina' x 'Forno' RILs. *QLr.sfr-1BS* (subsequently referred to as *Lr75*) was identified as the strongest leaf rust resistance QTL that was not associated with LTN. *Lr75* showed an additive effect with four other QTLs: *Lr34* which is associated with LTN, two minor QTLs that were contributed by 'Arina' and *QLr.sfr-7BL*, a minor QTL contributed by 'Forno' that was not linked to LTN. Based on these results we selected *Lr75* and *QLr.sfr-7BL* as candidates for marker-assisted introgression into the susceptible cultivar 'Arina'. To validate this strategy, we first re-evaluated the effect of this gene combination based on the original phenotypic data of the RIL population developed by Schnurbusch et al. (2004). *Lr34* was also included for comparison. For this, we grouped a subset of 117 RILs into different classes based on the availability of unambiguous marker information for the target regions of the three QTLs (Fig. 2.1). These groups were made based on the following markers: *gwm604* and *gwm131* for *Lr75*; *cssfr1* and *cssfr2* for *Lr34* (Lagudah et al. 2009) and *ksuD2* and *gbxGb218* for *QLr.sfr-7BL*.

In agreement with Schnurbusch et al. (2004), *Lr34* showed the strongest leaf rust resistance provided by a single QTL. *QLr.sfr-7BL* alone did not result in increased leaf rust resistance and *Lr75* alone showed a leaf rust resistance level that was weaker than *Lr34*. However, the combination of *Lr75* and *QLr.sfr-7BL* resulted in leaf rust resistance levels similar to *Lr34*, confirming that *Lr75* and *QLr.sfr-7BL* are additive. Interestingly, the combination of *Lr75* and *Lr34* did not improve leaf rust resistance compared to the presence of single genes, suggesting that *Lr75* and *Lr34* are not additive (Fig. 2.1). All gene combinations with *Lr34* were associated with LTN whereas the ones without *Lr34* were not. Based on the re-evaluation of these original RIL data we expected that the combination of *Lr75* and *QLr.sfr-7BL* would result in high leaf rust resistance levels similar to *Lr34* but without LTN. *Lr75* and *QLr.sfr-7BL* were therefore co-introduced into the genetic background of 'Arina' through marker-assisted backcrossing.

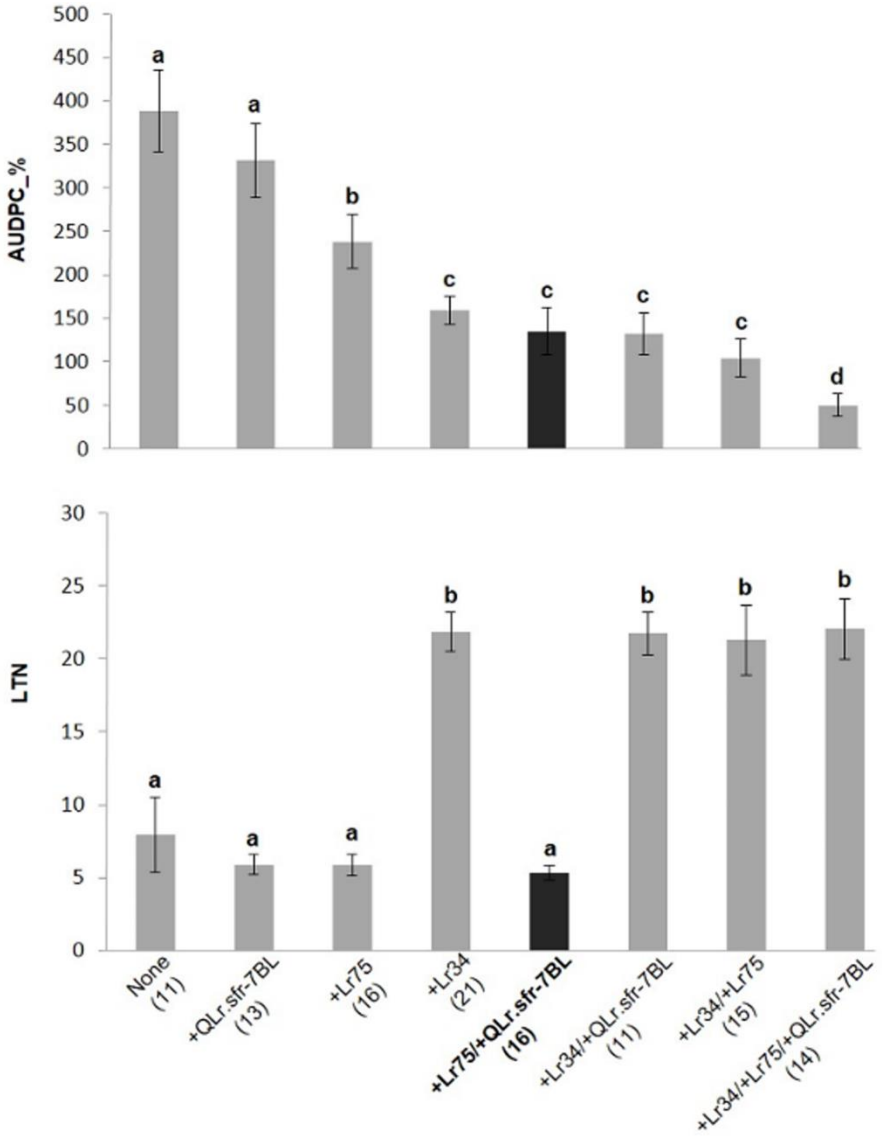


Fig. 2.1 Phenotypic effects of different leaf rust resistance QTL combinations. The phenotypic data of Schnurbusch et al. (2004) were evaluated for AUDPC_% (top graph) and LTN (bottom graph) on groups of RIL lines. The group of RILs with both *Lr75* and *QLr.sfr-7BL* is highlighted in black. Numbers in brackets indicate the number of RIL lines present in each class. Letters indicate lines with equivalent resistance levels (P>0.05, Student's *t*-test) and error bars represent standard errors of the mean.

2.4.2 Development and evaluation of a near isogenic line, 'Arina*LrFor*', carrying *Lr75* and *QLr.sfr7BL*

In order to introgress both *Lr75* and *QLr.sfr-7BL* into the susceptible cv. 'Arina', 101 BC_2F_5 back-cross lines were generated as described by Krattinger et al. (2009). The back-cross lines were screened with the flanking markers of the *Lr75* region (*barc128-gwm131*). Twelve of the 101 backcross lines were homozygous for the 'Forno' allele in the *Lr75* region. These twelve lines were further screened for presence of the *QLr.sfr-7BL* region (*gwm146-gwm344*) (Schnurbusch et al. 2004). In addition, these twelve lines were also screened for absence of *Lr34* with the diagnostic markers *cssfr1* and *cssfr2* (Lagudah et al. 2009). Out of these twelve lines, two lines were positive for *Lr34* and were therefore excluded. From the remaining 10 lines, one line, 'Arina*LrFor'*, was homozygous for

both *Lr75* and *QLr.sfr-7BL*. We also developed an 'Arina*Lr34*' back-cross line where the *Lr34* gene was introgressed into 'Arina' (BC_3F_7). The presence of *Lr34* in this line was confirmed with diagnostic markers, *cssfr1* and *cssfr2* (Lagudah et al. 2009).

The lines, 'Arina*LrFor*' and 'Arina*Lr34*' were evaluated together with 'Arina' and 'Forno' for leaf rust resistance in Switzerland and Australia (Table 2.1). In contrast to the results obtained in the RIL population, the leaf rust resistance of 'Arina*LrFor*' was stronger than 'Arina*Lr34*' in Switzerland (Table 2.1, Fig. 2.2) during three consecutive field seasons. 'Arina*LrFor*' displayed a slow-rusting response with a final leaf area coverage ranging from 14-36% in comparison to 'Arina*Lr34*' which had final leaf area coverage of 26-56%. 'Forno' displayed a near-immune response which is due to the combination of *Lr75, Lr34, QLr.sfr-7BL* and several minor QTLs. The same lines were also tested for leaf rust resistance in Australia. Similar to the results obtained for the Swiss environment, 'Arina*LrFor*' showed increased leaf rust resistance in the field in Australia (Table 2.1; Appendix D). In contrast to Switzerland however, the resistance of 'Arina*Lr34*' was stronger than 'Arina*LrFor*'. In summary, 'Arina*LrFor*' showed high levels of partial leaf rust resistance under Swiss and Australian environments. The results also indicate that the level of resistance provided by the gene combination in 'Arina*LrFor*' is environmental dependent.

Table 2.1 Adult plant field leaf rust response of Arina*LrFor*, Arina, Forno and Arina*Lr34* was tested at Agroscope Reckenholz, Zurich, Switzerland and Cobbitty, Australia

Genotype	Infection type (IT) (0-4 scale)	Rust severity (%)			
	Reckenholz, Switzerland	Reckenholz, Switzerland			Cobbitty, Australia
	2012	2013	2014	2015	2014
Arina	4	70	67	97	60-70
Forno	1	0	0	0	0
Arina*LrFor*	2	15	14	36	30
Arina*Lr34*	3.5	-[a]	26	56	20

[a] Arina*Lr34* was not included in the 2013 field trial

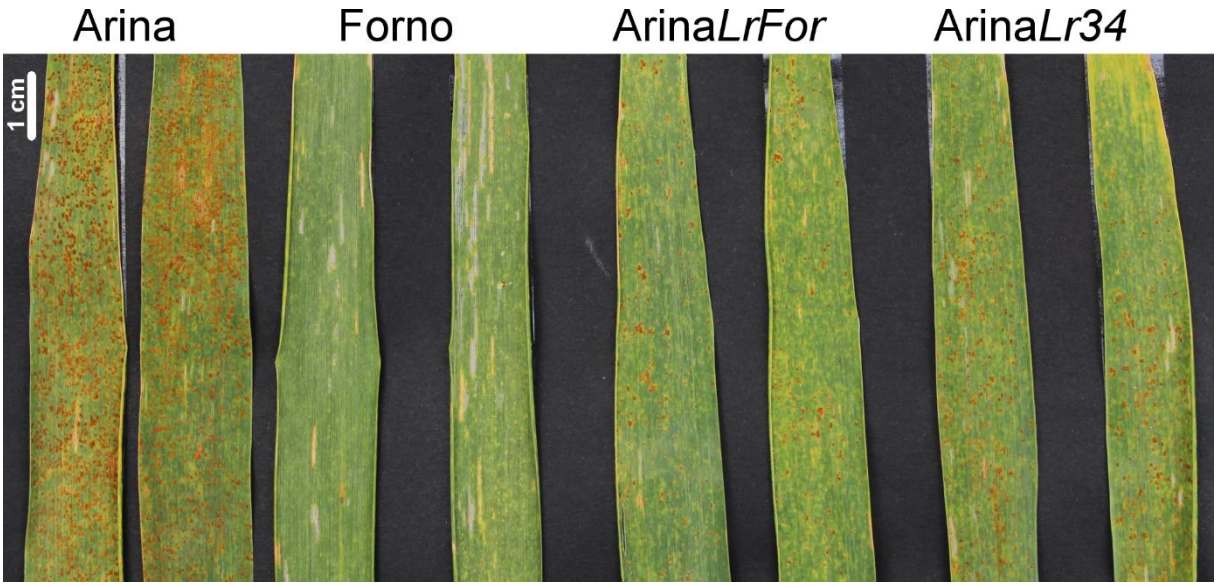


Fig. 2.2 Leaf rust infection on the flag leaves of Arina, Forno, Arina*LrFor* and Arina*Lr34*. Photographs were taken on field-infected plants in Switzerland in 2015.

Slow-rusting resistance genes are generally associated with a longer latency period, lower uredinial density and smaller uredinial size (Das et al. 1993). The line 'Arina*LrFor*' has a slow-rusting phenotype as shown in Fig. 2.2. Therefore, we quantified pustule density on the flag leaves of 'Arina', 'Forno', 'Arina*LrFor*' and 'Arina*Lr34*' on field-infected plants at three different time points when plants are between Zadkos growth stage of 55-69 (Zadoks et al. 1974) (Table 2.2). 'Forno' displayed significantly lower numbers of pustules than the other three lines and, in agreement with the near-immune phenotype, no significant increase in the pustule number was observed over time (84, 87 and 93 days after infection). On the other hand, 'Arina' showed a constant increase in the number of pustules from day 84 to day 93 after infection. Both 'Arina*LrFor*' and 'Arina*Lr34*' showed an intermediate response with a slower increase of pustule density observed in 'Arina*LrFor*' compared to 'Arina*Lr34*'.

Table 2.2 Pustule density was measured on the flag leaves at three different time points (84, 87 and 93 days after infection, dai) on 'Arina', 'Forno', 'Arina*LrFor*' and 'Arina*Lr34*' during the year 2014. Letters indicate lines with similar infection levels (P > 0.05, Student's *t*-test)

Genotype	Number of pustules/cm^2		
	84dai	*87dai*	*93dai*
Arina	5.07±5.12^{ab}	16.71±12.72^{c}	60.41±33.29^{e}
Forno	0.18±0.21^{a}	0.59±0.88^{a}	1.31±1.94^{a}
Arina*LrFor*	1.18±0.67^{a}	3.24±2.73^{a}	14.03±9.51^{c}
Arina*Lr34*	4.04±4.48^{a}	11.39±10.83^{bc}	28.31±15.23^{d}

2.4.3 'Arina*LrFor*' shows race-specific resistance at seedling stage

Partial adult plant resistance genes often only work at the adult plant stage and do not confer seedling resistance. In order to test the response of 'Arina*LrFor*' at seedling stage, we infected this line along with 'Arina', 'Forno' and 'Arina*Lr34*' in the greenhouse with seven Swiss leaf rust isolates that were also included in the artificial field infections. For the three isolates 91047, 95219 and 95001 'Arina*LrFor*' was as susceptible as 'Arina' and showed moderate to high infection types (IT=3 – 4) (Fig. 2.3a). Interestingly, 'Arina*LrFor*' displayed a more resistant response with infection types ranging from ; to 2 after infection with isolates 90035, 96002, 93012 and 96209 (Fig. 2.3b). This was surprising because 'Forno' showed a susceptible infection type to all isolates, despite the presence of *Lr75* and *QLr.sfr-7BL* in this cultivar. These results indicate that *Lr75* or *QLr.sfr-7BL* may interact and enhance the effect of a seedling resistance gene present in the genetic background of 'Arina' but absent in 'Forno'. Alternatively, the resistance in 'Arina*LrFor*' might be due to the removal of a suppressor gene present in 'Forno' which suppresses the activity of *Lr75* and *QLr.sfr-7BL* at the seedling stage.

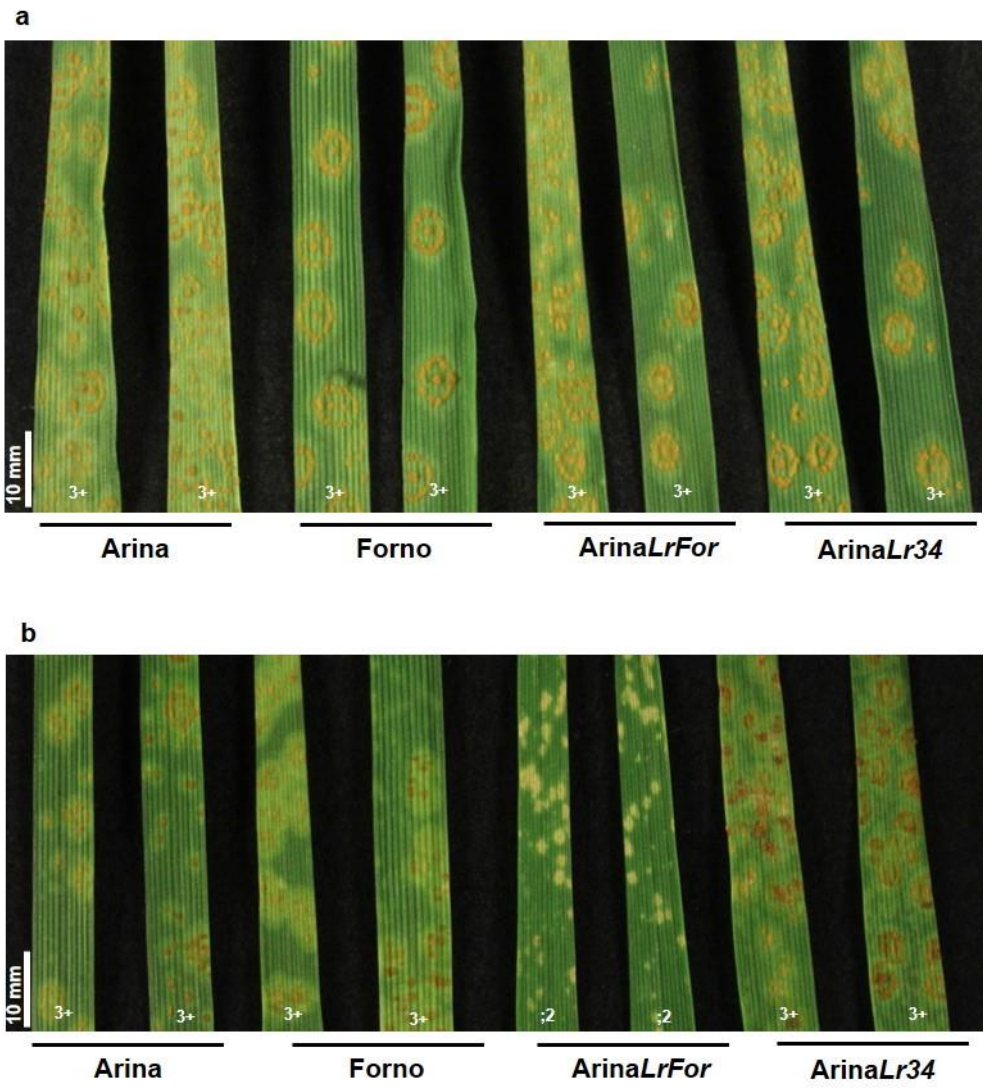


Fig. 2.3 Seedling infection assay on 'Arina', 'Forno', 'Arina*LrFor'* and 'Arina*Lr34'* with isolate (a) 91047 and (b) 90035. Infection type response was scored based on a 0-4 scale (Roelfs et al. 1992).

2.4.4 Genetic mapping of *Lr75*

Based on the QTL study of Schnurbusch et al. (2004) and our results, *Lr75* can be considered as a major, partial leaf rust resistance QTL. We therefore decided to further narrow down the *Lr75* interval. In the RIL population, *Lr75* was mapped on chromosome 1BS close to marker *gwm604* in an interval of 16 cM (Schnurbusch et al. 2004). To define an *Lr75* target interval we tested 63 publically available, 1BS-specific SSR markers (http://wheat.pw.usda.gov/GG2/index.html). Of the 63 SSRs, nine were 1B-specific and polymorphic (*barc128, cwem6c, cfa2158, gpw4069, wmc230, gwm11, gwm18, wmc277* and *wmc156)* between 'Arina' and 'Forno' and were added to the genetic map of the 'Arina' x 'Forno' RIL population (Schnurbusch et al. 2004). This resulted in the establishment of an 8 cM target region spanning the *Lr75* gene with *wmc230* and *gwm18* as the distal and proximal flanking markers, respectively. For precise genetic mapping and phenotypic analysis of *Lr75*, a near isogenic line (NIL) population was developed by further back-crossing 'Arina*LrFor*' with the susceptible

cultivar ‘Arina’. The resulting population consisted of 2,067 F_2 individuals. Out of these, 234 lines showed recombination between the two flanking markers *wmc230* and *gwm18*. These recombinants were further screened with two *QLr.sfr-7BL*-associated SSR markers (*gwm344* and *gwm146*). Only recombinants without the *QLr.sfr-7BL* QTL were selected for further mapping in order to avoid interference from the 7BL QTL during phenotyping. 65 BC_3F_2 recombinants were homozygous for the ‘Arina’ allele at the second locus on chromosome 7BL. These recombinants were phenotyped as F_3, F_4 and F_5 rows in the field in 2013 by using the same mixture of 16 Swiss leaf rust isolates as mentioned in material and methods.

To further saturate the 8 cM target region between *wmc230* and *gwm18* with additional markers, we exploited the flow-sorted Illumina survey sequences of chromosome 1BS of wheat cv. ‘Chinese Spring’ (http://wheat-urgi.versailles.inra.fr/Seq-Repository), the 1BS physical map generated by Raats et al. (2013) and information obtained by comparative genetics from a 1BS reference zipper based on synteny information of *Brachypodium*, rice and sorghum. The 1BS reference zipper was constructed in a similar manner as described by Breen et al. (2013). The gene-containing 1BS wheat Illumina sequences were physically anchored to BAC end sequences of wheat chromosome 1BS. These Illumina sequence contigs were further anchored to the reference zipper. The sequences of the flanking markers *wmc230* and *gwm18* were aligned against the integrated model of Illumina sequence contigs and reference zipper and the target region was defined. Then, we searched the Illumina sequences for microsatellite motifs within this target region and designed primers flanking the repeat motifs. Using this strategy, 98 SSR primers were designed (Appendix C), out of which 8 were polymorphic between the parents. Of these 8 primers, six (*swm271*, *swm275*, *swm276*, *swm278*, *swm281* and *swm294*) were mapped in the target interval in the BC_3F_2 fine mapping population (Fig. 2.4a). The other two markers, *swm216* and *swm247* were mapped at a distance of 0.17 and 0.32 cM proximal to *gwm18*, respectively. The addition of the 8 new SSR markers placed *Lr75* between the distal marker *gwm604* and proximal marker *swm271* at a distance of 1.6 and 2.7 cM, respectively (Fig. 2.4a).

2.4.5 Deletion bin mapping

In order to physically map *Lr75*, we used the cytogenetic stocks of chromosome 1B of wheat cv. ‘Chinese Spring’. Six deletion lines for the short arm (1BS4-sat-0.52, 1BS18-sat-0.50, 1BS2-sat-1.06, 1BS9-0.84, 1BS10-0.50 and 1BS1-0.35) and five deletion lines for the long arm (1BL11-0.23, 1BL6-0.32, 1BL1-0.47, 1BL2-0.69 and 1BL3-0.85) were used (Fig. 2.4b). Two ditelosomic lines DT1BS where the 1BL arm is missing and DT1BL where the 1BS arm is absent were also used. Marker *wmc230* amplified on none of the 1BS deletion lines but amplified in all lines with a deletion on the long arm. The marker *swm271* on the other hand did not amplify on deletion lines 1BS1-0.35 and 1BS10-0.50 (Fig. 2.4c, d). Hence, the use of the cytogenetic stocks physically placed the markers

wmc230 and *swm271* and the gene towards the distal end of chromosome 1BS (Fig. 2.4). (Singh et al. 2013b) mapped *Lr71* close to the centromere in the deletion bins 1BS10-0.50 and 1BL6-0.32, respectively (Fig. 2.4). Thus, in addition to the genetic information, physical deletion bin mapping also separates the *Lr75* gene from *Lr71*. Our results therefore prove that *Lr75* is a novel gene that maps in the distal region of the short arm of chromosome 1B.

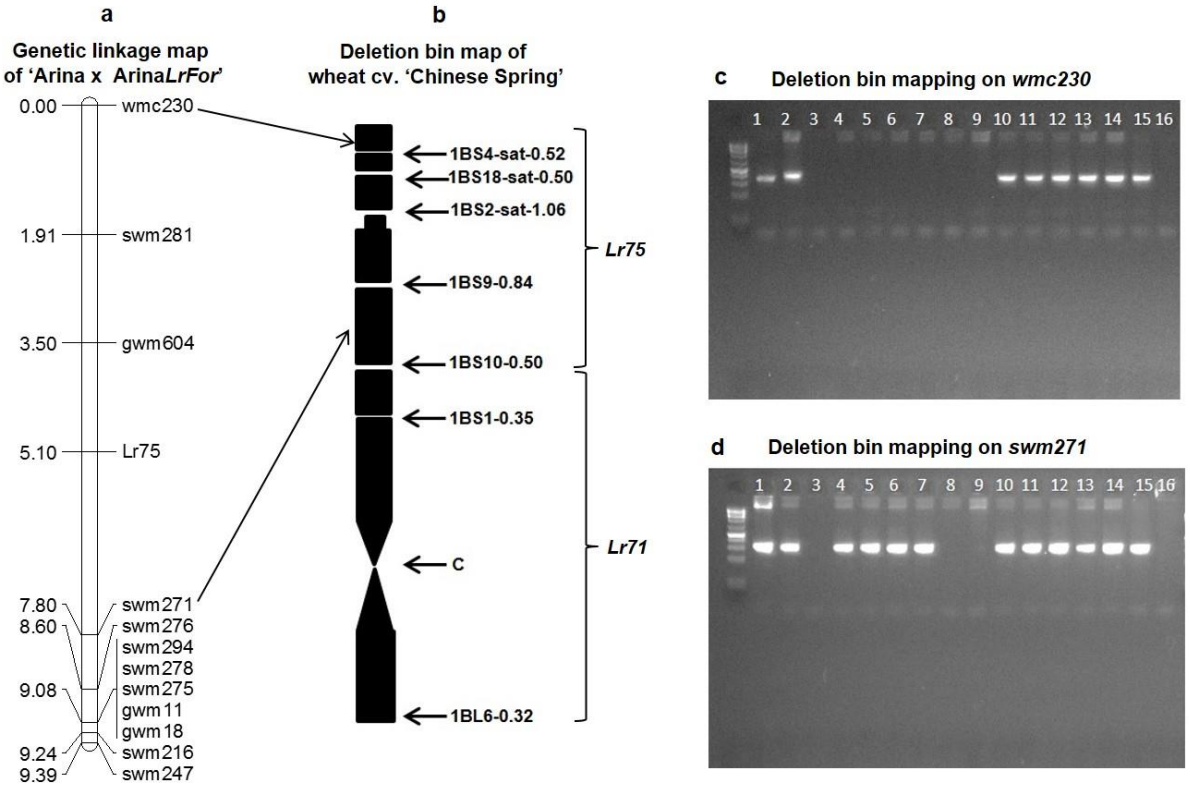


Fig. 2.4 Genetic linkage and physical deletion bin mapping of leaf rust resistance gene *Lr75*. a Genetic linkage map of the short arm of chromosome 1B of the 'Arina' x 'Arina*LrFor*' mapping population. Marker positions are shown in cM on the left side of the linkage map. b Deletion bin map of the short arm of chromosome 1B of wheat cv. 'Chinese Spring'. Physical bin localization of the markers *wmc230* and *swm271* is shown by arrowheads. c Deletion bin mapping of *wmc230*. d Deletion bin mapping of *swm271* on Chinese Spring, Arina*LrFor*, water, 1BS4-sat-0.52, 1BS18-sat-0.50, 1BS2-sat-1.06, 1BS9-0.84, 1BS10-0.50, 1BS1-0.35, 1BL11-0.23, 1BL6-0.32, 1BL1-0.47, 1BL2-0.69, 1BL3-0.85, DT1BS, DT1BL (*lanes 1-16*).

2.5 Discussion

Leaf rust resistance in the Swiss winter wheat cv. 'Forno' has been durable since 1986. The near immune response of 'Forno' against leaf rust is due to the combination of several major and minor leaf rust resistance QTLs (Schnurbusch et al. 2004). Resistance in 'Forno' is associated with the morphological marker leaf tip necrosis (LTN). In this study we introgressed two QTLs, *Lr75* and *QLr.sfr-7BL* that are not associated with LTN from 'Forno' into the leaf rust susceptible Swiss winter wheat cv. 'Arina'. Marker-assisted introgression of these two QTLs resulted in high levels of partial adult plant leaf rust resistance in 'Arina' in the field. Further, genetic mapping of *Lr75* with SSR markers placed this gene towards the distal end of chromosome 1BS. The only reported leaf rust resistance gene present on chromosome 1BS in the close proximity of *Lr75* is *Lr71* (Singh et al. 2013b). *Lr75* can be distinguished from *Lr71* by the marker *gwm18* which was reported to be the distal flanking marker of *Lr71* (Singh et al. 2013b), whereas *gwm18* mapped proximal to the *Lr75* gene. In addition, deletion bin mapping also physically separates *Lr75* from *Lr71*. Use of deletion lines on the SSR markers *wmc230* and *swm271* mapped *Lr75* towards the distal end whereas, deletion bin mapping mapped *Lr71* towards the centromere as reported by Singh et al. (2013b). Hence, both genetic and physical mapping of *Lr75* on chromosome 1BS with SSR markers showed that it is a novel gene as no other leaf rust resistance gene has been described in the target region of *Lr75*.

2.5.1 CIMMYT wheat germplasm as the major source of slow-rusting gene

Breeding for slow-rusting resistance has been a major objective of the CIMMYT rust improvement programme. CIMMYT wheat varieties are widely used throughout the world and get exposed to a wide range of rust races under environments that favour high pathogen pressure. The rapid evolution of rust pathogens results in the breakdown of race-specific resistance genes and gene combinations. A good example is the stem rust resistance gene *Sr31*, which has been widely used in wheat lines developed from CIMMYT programs. The resistance has been broken by the emergence of *Sr31*-virulent races referred to as Ug99. In order to increase and maintain rust resistance diversity in CIMMYT wheat germplasm, it became crucial for wheat breeders to establish a strategy for breeding of durable slow-rusting resistance in wheat. The strategy of slow-rusting resistance has been proposed by Caldwell (1968) and has been widely adopted in CIMMYT bread wheat germplasm improvement for more than 50 years. Singh and Rajaram (1991) showed that durable leaf rust resistance in CIMMYT lines is conferred by slow-rusting genes that have small to intermediate effects. Cultivars such as 'Frontana', 'Pavon76', 'Parula', 'Trap' or 'Mango' have been released during the past decades using CIMMYT bread wheat germplasm which show immune resistance response to wheat leaf rust by the additive effect of 3-4 slow-rusting genes (Singh and Rajaram 1991; Singh et al. 1998; Singh et al. 2005). The main APR genes identified in these cultivars are *Lr34/Yr18/Sr57/Pm38*, *Lr46/Yr29/Pm39* and *Lr68*. There are also other sources of slow-rusting resistance genes besides

CIMMYT germplasm. For example, the gene *Lr67/Yr46/Sr55/Pm46* has been identified in the common wheat accession PI250413 which was collected from Pakistan (Dyck and Samborski 1979; Moore et al. 2015).

All the above mentioned adult plant slow-rusting genes (*Lr34/Yr18/Sr57/Pm38*, *Lr46/Yr29/Pm39*, *Lr67/Yr46/Sr55/Pm46* and *Lr68*) are associated with leaf tip necrosis (LTN) which is considered undesirable in European wheat breeding programme (Singh et al. 1998; Krattinger et al. 2009; Hiebert et al. 2010; Herrera-Foessel et al. 2012). Due to this reason, the wheat cultivars possessing these genes have not been widely grown and accepted in European wheat breeding. It has therefore become essential to identify additional sources of durable rust resistance in European wheat germplasm without LTN. In our study we described yet another slow-rusting QTL, *Lr75* present on wheat chromosome 1BS. *Lr75* has shown to provide an additive effect when combined with another slow-rusting QTL, *QLr.sfr-7BL*. Both these QTLs are present in Swiss winter wheat cv. 'Forno' and are not associated with LTN. Another example of a non-LTN broad-spectrum APR gene is *Lr22a* which was introgressed from an *Aegilops tauschii* accession into cultivated wheat (Hiebert et al. 2007).

2.5.2 Breeding for slow-rusting resistance in European wheat germplasm

Very little information is available about leaf rust APRs in European wheat breeding material. Only a few studies looked at the existence of APR genes in ~100 European wheat lines (Winzeler et al. 2000; Park et al. 2001; Pathan and Park 2006). All studies reported the frequent occurrence of the *Lr13* APR gene in European wheat cultivars. Winzeler et al. (2000), detected varying levels of resistance shown by the cultivars carrying *Lr13* across Europe which indicates that virulence for *Lr13* exist. 'Arina' for example is known to possess *Lr13* but is highly susceptible to leaf rust throughout Europe (Pathan and Park 2006).

To our knowledge, only four studies reported on the identification of leaf rust APRs on chromosome 1BS in European wheat breeding material (Messmer et al. 2000; Schnurbusch et al. 2004; Singh et al. 2009; Buerstmayr et al. 2014). Other than 'Forno', cultivars 'Beaver' and 'Capo' possess 1BS QTLs. 'Beaver' has the 1BL/1RS translocation and the QTL identified in 'Beaver' can therefore not be *Lr75*. The QTL in 'Capo', *QLr.ifa-1B* mapped close to the centromere but since no detailed study has been available on this QTL, it is not clear whether the genomic location of *Lr75* and *QLr.ifa-1B* is identical. So far, 'Forno' seems to be the only source of *Lr75* and interestingly this gene has not been described in any other European wheat cultivar. However, apart from European wheat lines, two CIMMYT wheat lines, 'Pastor' and 'Parula' also possess a leaf rust resistance QTL on chromosome 1BS (William et al. 1997; Rosewarne et al. 2012). According to the study conducted by Singh and Rajaram (1992), the high level of resistance in 'Parula' is due to the combination of three

slow-rusting APR genes, *Lr34, Lr46* and *Lr68* plus some minor genes. William et al. (1997) by using a RIL population developed from a cross of resistant cv. 'Parula' and moderately susceptible cv. 'Siete Cerros' identified a minor QTL on chromosome 1BS in 'Parula' which explained a phenotypic variance of 7-10%. Another QTL on chromosome 1BS was detected in cv. 'Pastor' by Rosewarne et al. (2012). They reported this QTL to be present in the same genomic region as *Lr75*. Both these QTLs have not been characterized in detail and their precise genetic location is not available. Therefore, it is impossible to conclude if these QTLs are *Lr75* or not.

2.5.3 Slow-rusting APR genes are influenced by environment

The knowledge of an environmental influence on resistance genes allows wheat breeders to deploy resistance gene combinations most effectively in different regions. Our study showed that the combination of *Lr75* and *QLr.sfr-7BL* provided partial resistance in Switzerland and Australia. Similarly, *Lr34* also showed partial resistance at the adult plant stage in Switzerland and Australia. However, the level of resistance shown by these genes varied at the two locations. In Australia, the resistance provided by *Lr34* was stronger than that provided by the gene combination of *Lr75* and *QLr.sfr-7BL*, whereas in Switzerland, *Lr34* alone showed a weaker resistance response. Similar findings have also been reported in the literature where the environment plays a role in modifying the resistance response of slow-rusting genes. Herrera-Foessel et al. (2012) compared the leaf rust resistance response of *Lr68*, *Lr34*, *Lr46* and *Lr67* during three crop seasons (2008-2009, 2009-2010, 2010-2011) in the field at Ciudad Obregon, Mexico. They reported that except for one crop season (2010-2011), the effect of *Lr68* was smaller as compared to *Lr34*, *Lr46* and *Lr67*. In 2010-2011 however, *Lr68* showed a stronger resistance response than *Lr46*. Similar results were observed by Lillemo et al. (2011) while studying the additive effect of three APR genes, *Lr34/Yr18/Sr57/Pm38, Lr46/Yr29/Pm39* and *Lr68* in 'Avocet-YrA x Parula' F_6 RIL mapping populations across nine different environments. In agreement with Herrera-Foessel et al. (2012), they also observed a smaller resistance response of *Lr68* than *Lr34* and *Lr46* in Mexico. On the other hand, a stronger resistance response of *Lr68* as compared to *Lr34* was seen in Argentina and Uruguay (Lillemo et al. 2011). Interestingly, the combination of *Lr68* and *Lr34* showed stronger resistance than either gene alone in all the tested environments which suggest an additive effect of these two genes. In contrast, Silva et al. (2015) while studying the effect and interaction of *Lr68, Lr34* and *Sr2* genes in two wheat populations derived from 'Parula' at sites in Uruguay did not observe an additive effect of the combination of *Lr68* and *Lr34*. Instead, the effect of the combination of *Lr68* and *Lr34* was comparable to the effect of *Lr68* alone. These studies clearly showed that resistance gene combinations do not necessarily behave in the same manner in all environments and their resistance response can vary in different environments which might be due to variations in temperatures or other environmental conditions. Stem rust resistance gene, *Sr2* is known to be tightly linked to seedling resistance gene *Lr27* (Mago et al. 2011) and *Lr27*

was also reported to be responsible for reducing leaf rust severity (Bariana et al. 2007). Silva et al. (2015) studied the effect of the *Sr2* gene in reducing leaf rust in Uruguay. They observed that the stem rust resistance gene *Sr2* does not have any effect on leaf rust resistance when present alone but a significant increase in resistance level was seen when *Sr2* was present in combination with *Lr68*. From their study it was clear that *Sr2* alone is not strong enough to provide resistance and rather it enhances the effect of *Lr68*. These results are in agreement to what we observed in the RIL population, where *QLr.sfr-7BL* individually has no effect on leaf rust resistance whereas it increased the resistance response to a significant level when combined with either *Lr75* or *Lr34*. Therefore, it is essential to study the effect of different APR genes either singly or in combinations in different environments so that effective gene combinations can be deployed in the corresponding target environment.

2.5.4 *QLr.sfr-7BL* might be the leaf rust resistance gene *Lr14a*

In a survey of wheat leaf rust in western Europe, a leaf rust resistance gene, *Lr14a* was found to be present in cv. 'Forno' (Park et al. 2001). This gene was mapped to the distal end of chromosome 7BL in a wheat consensus map (Gale et al. 1995). A major leaf rust resistance QTL, (*QLr.ubo-7B.2*) was also identified by Maccaferri et al. (2008) on chromosome 7BL within an 8.2 cM region in durum wheat cv. 'Creso'. Their study also postulated that this QTL is effective at both seedling and adult plant stages. The two microsatellite markers *gwm146* and *gwm344* were reported to be tightly linked to this QTL on chromosome 7BL. The same SSR markers were also reported to be closely linked to another gene, *LrLla* which was more precisely mapped on chromosome 7BL in a population of 98 F_3 lines derived from Chilean durum cv. Llareta INIA by Herrera-Foessel et al. (2008) . They postulated this gene to be *Lr14a* based on the resistance response and chromosomal location. In addition, (Singh et al. 2013a) also identified a major QTL for leaf rust resistance on chromosome 7BL close to marker *gwm146* in French durum wheat cv. 'Sachem'. They reported that this QTL is effective at both seedling and adult plant stages when tested in different environments in Mexico. The APR gene *Lr68* was also mapped on chromosome 7BL in the same genomic region as *QLr.sfr-7BL* by Herrera-Foessel et al. (2012). Like all the mentioned QTLs, *QLr.sfr-7BL* identified from 'Forno' also shared the same genomic region on chromosome 7BL close to the markers, *gwm344* and *gwm146*. Therefore, it is possible that this QTL is actually the *Lr14a* gene. *Lr14a* gene was transferred from emmer wheat to modern bread wheat cv. 'Hope' and H-44. Park et al. (2001) reported the frequent occurrence of *Lr14a* in Europe and that 33.5% of the area in France has been occupied by *Lr14a* alone or in combination with *Lr13*. Virulence against *Lr14a* was reported in Europe by Goyeau et al. (2010) and Park et al. (2001). In addition, all the seven leaf rust isolates used in the seedling assay (Fig. 2.2) are virulent on 'Arina*LrFor*' (Appendix B). Hence, race-specificity of 'Arina*LrFor*' at the seedling stage cannot be due to *Lr14a*. Therefore, it has become essential to carefully deploy this gene in wheat resistance

breeding Several studies detected QTLs on chromosome 7BL in this genomic region, but further detailed study will be required to resolve the relationship between *QLr.sfr-7BL* and these QTLs.

Chapter 3

Marker development based on flow-sorted sequenced chromosome 1B of wheat cultivars Arina and Forno

3.1 Summary

Genetic mapping of agronomically important genes in the wheat genome has been a challenging task due to its polyploid nature, repetitive content and the low level of polymorphism among wheat cultivars. This creates difficulties in the development of genome-specific, highly reproducible and polymorphic markers. The advancement in next generation sequencing technologies (NGS) accelerates the development of genome-specific polymorphic markers and facilitates genetic mapping and map-based cloning in wheat. We used different strategies for marker development to construct a high-density genetic map of wheat chromosome 1B. Two different mapping populations consisting of recombinant inbred lines (RIL) and near-isogenic lines (NIL), respectively were used in this study. Chromosome 1B of the two wheat cultivars 'Arina' and 'Forno' was flow-sorted which resulted in a 25% purity, and they were sequenced to 16-fold coverage by Illumina sequencing. Because of the mixture of sequences from other chromosomes in this enriched fraction, the syntenic information from *Brachypodium* chromosome 2 was used to identify the gene-associated sequence contigs of the two wheat cultivars. Using the predicted gene annotation of the syntenic region of *Brachypodium* chromosome 2, single nucleotide polymorphism (SNP) markers were developed from sequences located close to coding regions. As an alternative approach for SNP marker development, we used the 90K SNP array to identify SNPs between the two wheat cultivars. The third approach for SNP marker development was the use of BAC end sequence (BES) information of chromosome 1BS of wheat cv. 'Chinese Spring'. Using all these approaches, we developed in total 149 SNP markers, 142 from flow-sorted, sequenced chromosomes, four from the 90K SNP array and three from BES of chromosome 1BS. The addition of SNP markers in the NIL population reduced the *Lr75* target region from 4.3 cM to 1.2 cM. Thus, in our study we used different strategies that permit fast and efficient development of genome-specific, gene-associated SNP markers in different wheat cultivars which can be useful for high resolution genetic mapping and map-based cloning.

3.2 Introduction

Allohexaploid bread wheat (*Triticum aestivum* L. 2n=6x=42) is one of the most important food crops of the world. Due to the presence of three A, B and D genomes, many wheat genes occur as three homeologs. In addition, more than 80% of wheat genome consists of repetitive elements (Flavell et al. 1977; Hollister and Gaut 2009) resulting in a total genome size of 17 Gb and complicating genetic mapping in wheat. In order to overcome these limitations, several efforts have been undertaken by the wheat genetics community to establish different technologies for the development of molecular markers such as simple sequence repeats (SSRs), restriction fragment length polymorphism (RFLP) and single nucleotide polymorphisms (SNPs) (Paux et al. 2012).

To characterize and isolate genes of agronomical importance, the marker density in the target region surrounding the gene of interest has to be increased. This can be supported by using the syntenic information from the sequenced genomes of barley (Mayer et al. 2012), rice (*Oryza sativa* L., International Rice Genome Sequencing Project 2005), sorghum (*Sorghum bicolor* L., Paterson et al. 2009) and *Brachypodium* (*Brachypodium distachyon*, Vogel et al. 2010). These plants can be used as model organisms for grass crops with larger genome size such as wheat as their genomes are of small size and because of the conservation of gene order between different members of the grass family (Vogel et al. 2010). Among all the model grass genomes, *Brachypodium* has the smallest genome size (355 Mb) and its genome diverged from the one of wheat about 32-39 million years ago (MYA). Due to these reasons *Brachypodium* has been used for construction of genetic maps based on synteny as well as to predict the gene order in Triticeae (Wicker et al. 2009).

There has been a dramatic advancement in wheat genomics during the past few years due to the efforts undertaken by the International Wheat Genome Sequencing Consortium (IWGSC) and others in order to physically map and sequence the wheat genome. Isolation of individual chromosome and chromosome arms by cytometric flow-sorting has been a major breakthrough in order to reduce the complexity of wheat genome analysis (Safár et al. 2010). Using this strategy, chromosome specific BAC (bacterial artificial chromosome) libraries of the wheat cv. 'Chinese Spring' have been generated and used for physical map construction and sequencing of individual chromosomes and chromosome arms. Using this approach, a physical map of wheat chromosome 1BS of cv. 'Chinese Spring' was constructed by Raats et al. (2013). Very recently, sequencing of the isolated chromosome arms resulted in the production of a draft sequence of hexaploid bread wheat cv. 'Chinese Spring' by IWGSC (The International Wheat Genome Sequencing Consortium 2014).

Molecular markers have proved to be an efficient tool for genetic mapping and map-based cloning in different crop species. For a long time, the generation of polymorphic markers relied on a set of limited resources such as the data generated from EST, wheat BAC libraries as well as syntenic information from rice, sorghum and *Brachypodium.* Microsatellite or simple sequence repeat (SSR) markers replaced earlier technologies such as RAPD, RFLP and AFLP markers. With the increasing

possibilities for partial sequencing of genomes of different wheat cultivars, SSRs were more and more replaced by SNPs. Several thousand SNPs have been genotyped recently by using the survey sequencing data from different genomes (Chao et al. 2009; Akhunov et al. 2009; Allen et al. 2011). In combination with high-throughput genotyping, SNPs are used for high resolution mapping, marker-assisted breeding and genetic diversity studies (Agarwal et al. 2008; Zhao et al. 2011). SNP densities in modern wheat varieties as compared to other crops are very low due to the high genome similarity between modern wheat cultivars of approximately 99.9% (Ravel et al. 2006; Chao et al. 2009; Trick et al. 2012). Ching et al. (2002) observed an average SNP density of 1 SNP in 47 bp in 36 maize inbred lines. In eight diverse barley accessions Rostoks et al. (2005) observed a SNP density of one SNP every 200 bp. In wheat, Ravel et al. (2006) estimated the SNP frequency of 1 SNP in 330 bp across 26 cultivars representing a diverse wheat germplasm. The SNP density is further reduced in less diverse germplasm. For instance, Somers et al. (2003) used 12 wheat accessions and estimated the frequency of 1SNP/540 bp and Trick et al. (2012) in their study detected a SNP density of 1.8 SNPs per kb in wheat.

Two Swiss winter wheat cultivars, 'Arina' and 'Forno' were used as parents to generate a recombinant inbred line population (RIL) (Paillard et al. 2003). This population has been used in various QTL mapping studies to map genes for important traits. Schnurbusch et al. (2004) used this population for QTL mapping of leaf rust resistance loci on different wheat chromosomes and constructed genetic linkage maps for all chromosomes. The genetic linkage map of chromosome 1B spanned a length of 183 cM with 23 markers. Schnurbusch et al. (2004) identified a new QTL, *QLr.sfr-1BS* on the short arm of chromosome 1B. We refer to this QTL as *Lr75* hereafter. In order to fine map this QTL, we generated a fine mapping population by crossing a backcross line, 'Arina*LrFor*' with susceptible cv. 'Arina'. Details of the development of the fine mapping population (near isogenic line, NIL) are described in chapter 2 of this thesis. *Lr75* was mapped in an interval of 4.3 cM between the markers *gwm604* and *swm271* as described in chapter 2.

Here, we report on the different strategies used for marker development in order to enrich the genetic linkage maps of both RIL and NIL populations. We sequenced a flow-sorted fraction enriched for the 1B chromosome of cultivars, 'Arina' and 'Forno' and developed genome-specific gene-associated SNP markers from the sequences of 'Arina' and 'Forno'. We developed 142 SNP markers, of which 74 and 63 SNPs were mapped on RIL and NIL population, respectively. As an alternative strategy we also used the 90K SNP array (Wang et al. 2014) and mapped additional four SNP markers in both populations. Our analysis suggests a good conservation of gene order between chromosome 1BS of wheat and *Brachypodium* chromosome 2 with some minor re-arrangements and inversions.

3.3 Materials and Methods

3.3.1 Plant material

The parental genotypes 'Arina', 'Forno' and 'Arina*LrFor*' were selected for SNP analysis. Two different mapping populations were used for genotyping. The first population consisted of recombinant inbred lines (RIL) generated from the cross of two Swiss winter wheat cultivars, 'Arina' and 'Forno' as described in Paillard et al. (2003). The second population was derived from a cross of 'Arina' and resistant backcross line 'Arina*LrFor*' as described in chapter 2 of this thesis. DNA was extracted from the leaf tissue as described in (Stein et al. 2001).

3.3.2 Flow-sorting of chromosome 1B

Chromosome 1B of cultivars 'Arina' and 'Forno' was enriched by flow-sorting the liquid suspensions of mitotic chromosomes from root tips (Vrána et al. 2000). Chromosome 1B was sorted with a purity of 24% and 29% for 'Arina' and 'Forno', respectively, using the protocol as described in Kubaláková et al. (2002). DNA was amplified using Illustra GenomiPhi DNA amplification kit (GE Healthcare). Three independent amplifications were made for each cultivar and in total 17.16 µg and 16.05 µg DNA were obtained for 'Arina' and 'Forno', respectively.

3.3.3 Sequencing and assembly of chromosome 1B of 'Arina' and 'Forno'

Sequencing of chromosome 1B of both cultivars was done by GATC Biotech, Konstanz, Germany. For sequencing, 5 µg of DNA was used for each cultivar. Illumina HiSeq 2000 with paired end reads of 100 bp was used for sequencing. After quality trimming, 'clc_novo_assemble' command was used for *de novo* assembly of 100-bp-long paired end reads of 'Arina' and 'Forno'. To create the assembly, the default parameters of CLC Assembly Cell 3.22 software (CLC Bio, Aarhus, Denmark) were used.

3.3.4 Identification of SNPs for genotyping

'Arina' reads were mapped to the *de novo* sequence assembly of cv. 'Forno'. The function 'clc_ref_assemble_long' of CLC Assembly Cell was used for mapping of 'Arina' reads. By using the function 'find_variations' of the CLC Assembly Cell, potential SNPs were identified. For each base pair position in targeted contigs, read coverage was extracted by using the function 'assembly_info' of the CLC Assembly Cell. For genotyping the parents and a subset of 24 BC_3F_2 NIL lines using 90K SNP array developed by Wang et al. (2014), 10 µL of DNA with a concentration of 50 ng/µL was sent to Trait Genetics, GmBH, Germany.

3.3.5 Genetic map construction

To construct the genetic map of 144 selected SNPs, 169 and 65 lines of the respective RIL and NIL population were used. The genotyping was done by KBioscience (Herts, UK) using fluorescence based kompetitive allele-specific PCR (KASPar) genotyping technology (http://www.lgcgroup.com). The genetic map of the RIL population was constructed using mapping software MapMaker (Lander et

al. 1987). To calculate the distance between the markers, the Kosambi function was used (Kosambi 1944). Recombination frequency between the markers was calculated for construction of the genetic map of the NIL population.

3.3.6 Marker development using high resolution melting (HRM) analysis

For HRM amplification, forward and reverse primers spanning the putative SNPs between the parents were designed using Primer 3.0 from NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The expected product size amplified by the primers was set in the range of 70-150 bp and the annealing temperature between 61-63°C. The PCR amplification on the parental lines and NIL population was done using the CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories AG, Switzerland). For amplification of HRM based primers the following protocol was used:

HRM standard protocol

2.4 µl	H_20	95°C for 2 min	
5 µl	HRM Precision Melt Supermix (Bio-Rad Laboratories AG, Switzerland)	95°C for 10 s 63°C for 30 s	} 40 cycles
0.3 µl	Forward primer 300 µM	95°C for 30 s	
0.3 µl	Reverse primer 300 µM	60°C for 1 min	
2 µl	Template DNA (13 ng/µl)	65.0 to 95.0°C, increment 0.2°C for 10 s	
10 µl	Total reaction volume		

3.4 Results

3.4.1 Use of flow cytometry sorting to generate a chromosome specific SNP map

3.4.1.1 Purification and sequencing of chromosome 1B of 'Arina' and 'Forno'

In order to reduce the complexity of wheat sequences and to eliminate the homeologous 1A and 1D sub-genomes, chromosome 1B of two parental winter wheat cultivars, 'Arina' and 'Forno' was purified by flow cytometry sorting. The flow-sorting resulted in a flow karyotype with three composite peaks each containing a mixture of several chromosomes. Chromosome 1B was present in peak III together with 8 other chromosomes (2A, 3A, 4A, 4B, 5A, 5B, 6B and 7A). Fortunately, no homeolog of chromosome 1B was present in this peak. The flow-sorting resulted in a purity of 24% and 29% of chromosome 1B for 'Arina' and 'Forno', respectively. The isolated 1B chromosome of both cultivars was sequenced by Illumina sequence technology. The numbers of reads with an average read length of 101 bp obtained by sequencing of the flow-sorted chromosomes were 536,380,252 and 563,436,688 for 'Arina' and 'Forno', respectively. *De novo* assemblies of 'Arina' and 'Forno' resulted in 1,306,238 and 1,519,864 contigs, respectively. To identify polymorphisms *de novo* sequence assembly of 'Forno' was chosen as reference on to which 'Arina' Illumina reads were mapped. Mapping resulted in 181,954,547 mapped reads.

3.4.1.2 Gene identification using *Brachypodium* syntenic information and SNP detection in gene-containing 'Forno' contigs

The syntenic information of *Brachypodium* was used in order to obtain the 1B-specific sequence contigs of both cultivars because of the mixture of chromosome 1B with other chromosome sequences. In addition, to identify the gene-containing 'Forno' contigs, the reference *de novo* sequence assembly of 'Forno' was compared against the *Brachypodium* coding sequence database using BLASTN. A total of 68,074 'Forno' contigs containing coding sequences were obtained which corresponds to approximately 4.5% of all 'Forno' contigs. This number consists of three different groups. The first group contained the 'Forno' contigs with a single gene hit in *Brachypodium*. We called such genes single-copy genes. Second, when two or more different 'Forno' contigs hit different parts of the same *Brachypodium* gene with little or no overlap, those contigs were considered as parts of the same gene. Third, when two or more different 'Forno' contigs consisted of homologs of the same *Brachypodium* gene with different sequence, they were called paralogous copies. By combining all three groups, we established a set of 'non-redundant' genes which consists of all *Brachypodium* genes for which the homologs were present in chromosome 1B 'Forno' contigs irrespective of the situation where several 'Forno' contigs hit the same *Brachypodium* gene. For instance, if a single *Brachypodium* gene has a hit with several 'Forno' contigs, that gene was counted only once in the non-redundant gene set. We detected 13,717 non-redundant genes on 68,074 'Forno' contigs.

Out of the 68,074 ‘Forno’ contigs, 17,304 contigs had hits with *Brachypodium* chromosome 2 genes. This number corresponds to what we expected because of the 25% purity of chromosome 1B of wheat due to the mixture with other chromosome sequences. Because of the chromosome fusion in *Brachypodium*, chromosome 2 of *Brachypodium* is syntenic to chromosomes 1 and 3 of wheat. Therefore, our main focus was on the contigs which had hits with *Brachypodium* chromosome 2 genes that are syntenic to wheat chromosome 1. The central region of *Brachypodium* chromosome 2 is syntenic to wheat chromosome 1: *Bd2g14080-Bd2g30400* corresponds to the long arm and *Bd2g30410-Bd2g40150* corresponds to the short arm. Out of 17,304 ‘Forno’ contigs which had hits with *Brachypodium* chromosome 2 genes, we selected only chromosome 1B-specific contigs based on synteny with *Brachypodium*. This resulted in 7,117 ‘Forno’ contigs which consisted of 1,353 non-redundant genes belonging to the 1B-syntenic region of *Brachypodium* chromosome 2. The remaining 10,287 contigs might belong to the wheat chromosome 3 syntenic region of *Brachypodium* because the peak which has chromosome 1B also contains chromosome 3A. *Brachypodium* chromosome 2 consists of a total of 2,191 genes in the syntenic region of wheat chromosome 1. This means that 62% of the expected *Brachypodium* chromosome 2 syntenic genes were present on ‘Forno’ sequence contigs.

Further, to detect polymorphisms between ‘Arina’ and ‘Forno’ sequences, we mapped ‘Arina’ Illumina reads on the reference *de novo* sequence assembly of ‘Forno’. CLC Assembly Cell was used to detect nucleotide variations between the two cultivars with a focus on the 7,117 non-redundant, gene-containing ‘Forno’ contigs. In order to exclude the contigs containing paralogous gene copies, only those ‘Forno’ contigs were selected which either hit a single gene or which did not overlap with other contigs which hit the same gene. To exclude sequencing errors an additional selection criteria was set among the ‘Forno’ contigs: All the low coverage SNPs i.e SNPs with coverage lower than 5 were eliminated from the analysis. Similarly, all polymorphic regions containing SNPs between ‘Arina’ and ‘Forno’ were eliminated from the analysis where the coverage of reads was less than 5 in ‘Arina’. This resulted in 616 ‘Forno’ contigs with 1,979 SNPs and 66 InDels that had gene hits with the syntenic regions of *Brachypodium* chromosome 2 (Fig. 3.1).

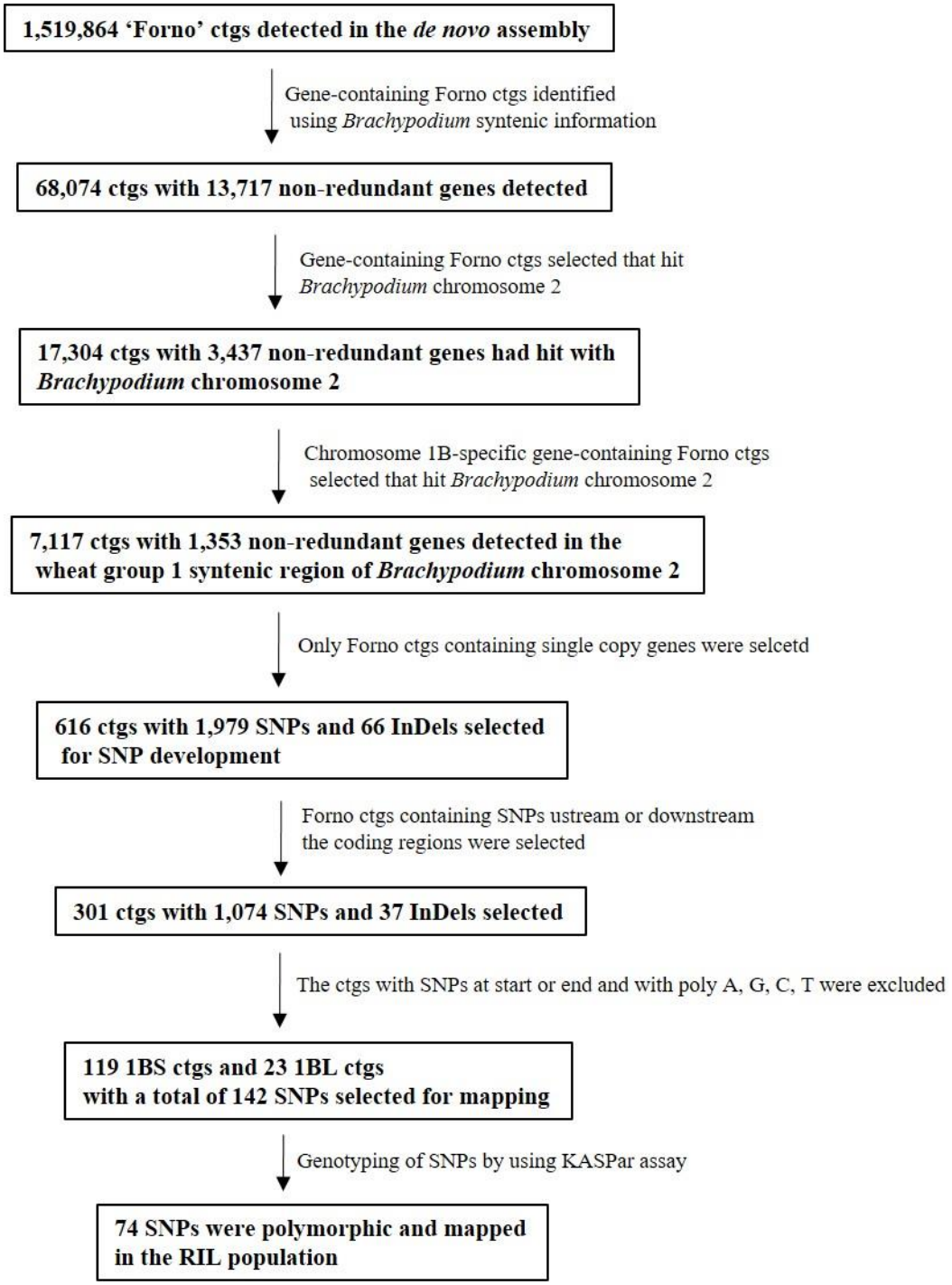


Fig. 3.1 Steps for SNP detection. Coding sequences on the contigs of 'Forno' assembly were detected using *Brachypodium* coding sequence database.

3.4.1.3 Development of a genetic map of chromosome 1B based on SNP markers

For the development of the SNP markers, out of 616 'Forno' contigs only those contigs were selected that have SNPs either upstream or downstream of the coding sequences and which were flanked by 100-200 bp of sequence on both sides. This SNP selection was done to possibly reduce amplification of the homeologous copies of the genes. This selection criterion resulted in the final number of 301 'Forno' contigs with 1,074 SNPs and 37 InDels. Out of 301 'Forno' contigs, 132 belong to the short arm and 169 belong to the long arm of chromosome 1B. Further, all contigs which had SNPs either at the beginning or end of the sequence or which were flanked by sequences with poly A, G, C or T were not considered for SNP development. For SNP development 119 1BS 'Forno' contigs and 23 1BL 'Forno' contigs were finally selected. The idea to select the 1BL contigs was to increase the marker resolution close to the centromere towards the long arm of the RIL genetic linkage map.

In total 142 SNPs, 119 from the short arm and 23 from the long arm of chromosome 1B were selected for mapping on 194 RILs and 65 NILs using KASPar assay. Out of 142 SNPs, 110 SNPs were amplified and the rest failed to amplify. Out of 110 SNPs, 74 were polymorphic and were mapped in the RIL population. The remaining 36 markers did not produce good genotypic data and had too many missing data. These markers were therefore excluded from the analysis. Among the 74 mapped markers, 63 were mapped in the NIL population (Fig. 3.2, Appendix E). The number of mapped markers in the NIL population is in agreement to what we expected because the complete short arm as well as some part of the long arm of the 'Forno' genome was introgressed into the 'Arina' background: The region from *barc128-gwm131* was introgressed into 'Arina*LrFor*' (Fig. 3.2, markers in pink and bold). In order to construct the genetic map of the RIL population, 25 RIL lines with too many missing genotypes were eliminated from further analysis and therefore, the data was analysed on 169 RILs. The initial size of the genetic map of chromosome 1B of the 'Arina x Forno' RIL population was 183 cM with 23 markers. After the integration of additional 74 SNPs and 17 SSR markers, the length of the RIL genetic map of chromosome 1B was increased to 184 cM. Because the length of the genetic map was increased by only 1 cM after the integration of 74 SNP markers, this clearly indicated that integration of the newly developed SNP markers in the RIL genetic map was of very good quality. The integration of 63 SNP markers on the chromosome 1B of the NIL genetic map increased the resolution of the genetic map with a map length of 18.3 cM. The SNP-based KASP markers were named based on the *Brachypodium* gene homologs present on the wheat contigs e.g. *Bradi2g38480* was named as marker *B2g38480*.

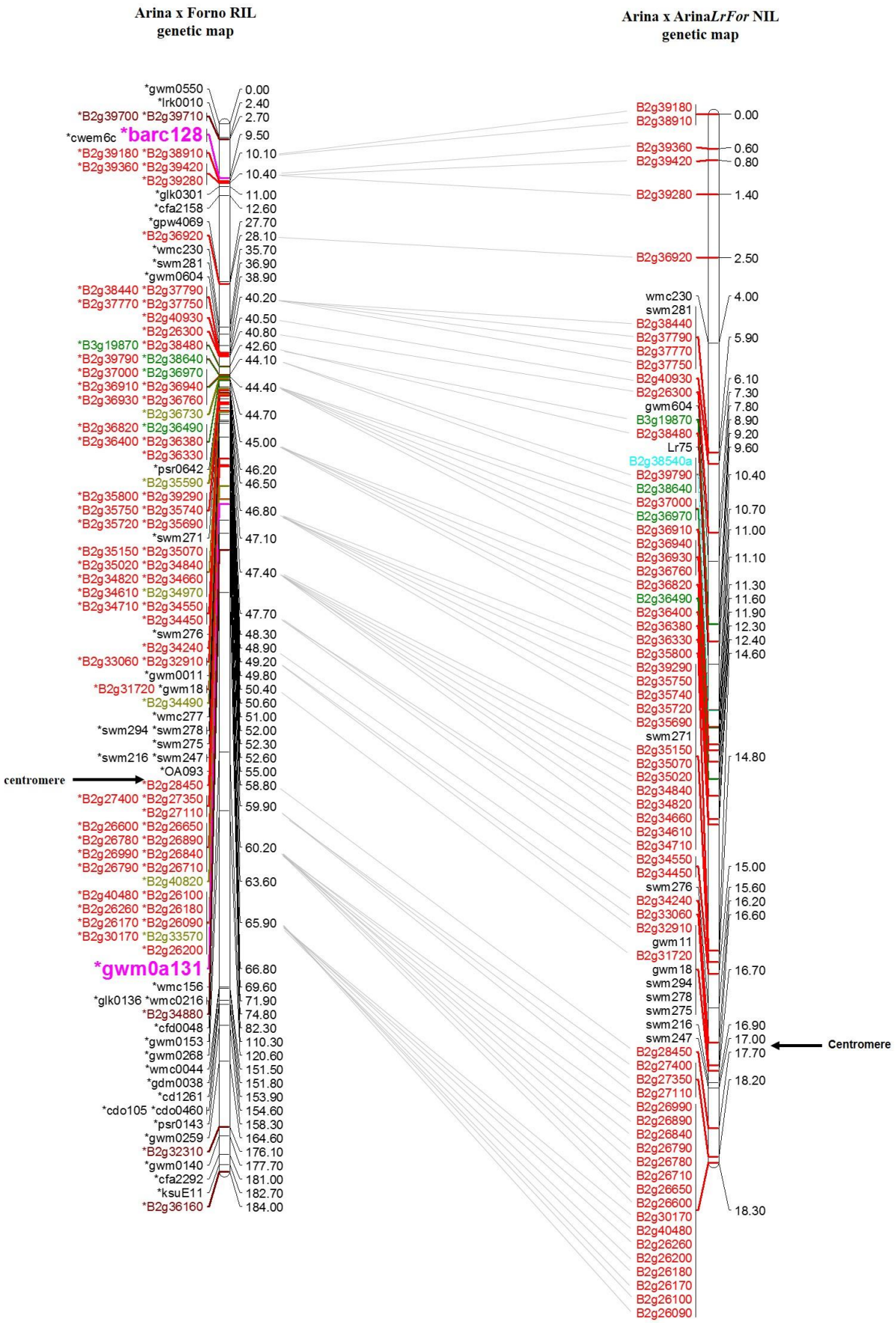

Arina x Forno RIL genetic map
Arina x ArinaLrFor NIL genetic map
*gwm0550 0.00
*lrk0010 2.40
*B2g39700 *B2g39710 2.70
*cwem6c *barc128 9.50
*B2g39180 *B2g38910 10.10
*B2g39360 *B2g39420 10.40
*B2g39280
*glk0301 11.00
*cfa2158 12.60
*gpw4069 27.70
*B2g36920 28.10
*wmc230 35.70
*swm281 36.90
*gwm0604 38.90
*B2g38440 *B2g37790 40.20
*B2g37770 *B2g37750
*B2g40930 40.50
*B2g26300 40.80
*B3g19870 *B2g38480 42.60
*B2g39790 *B2g38640 44.10
*B2g37000 *B2g36970
*B2g36910 *B2g36940 44.40
*B2g36930 *B2g36760
*B2g36730 44.70
*B2g36820 *B2g36490
*B2g36400 *B2g36380 45.00
*B2g36330
*psr0642 46.20
*B2g35590 46.50
*B2g35800 *B2g39290
*B2g35750 *B2g35740 46.80
*B2g35720 *B2g35690
*swm271 47.10
*B2g35150 *B2g35070
*B2g35020 *B2g34840
*B2g34820 *B2g34660 47.40
*B2g34610 *B2g34970
*B2g34710 *B2g34550 47.70
*B2g34450
*swm276 48.30
*B2g34240 48.90
*B2g33060 *B2g32910 49.20
*gwm0011 49.80
*B2g31720 *gwm18 50.40
*B2g34490 50.60
*wmc277 51.00
*swm294 *swm278 52.00
*swm275 52.30
*swm216 *swm247 52.60
*OA093 55.00
centromere
*B2g28450 58.80
*B2g27400 *B2g27350 59.90
*B2g27110
*B2g26600 *B2g26650
*B2g26780 *B2g26890 60.20
*B2g26990 *B2g26840
*B2g26790 *B2g26710
*B2g40820 63.60
*B2g40480 *B2g26100
*B2g26260 *B2g26180
*B2g26170 *B2g26090 65.90
*B2g30170 *B2g33570
*B2g26200
*gwm0a131 66.80
*wmc156 69.60
*glk0136 *wmc0216 71.90
*B2g34880 74.80
*cfd0048 82.30
*gwm0153 110.30
*gwm0268 120.60
*wmc0044 151.50
*gdm0038 151.80
*cd1261 153.90
*cdo105 *cdo0460 154.60
*psr0143 158.30
*gwm0259 164.60
*B2g32310 176.10
*gwm0140 177.70
*cfa2292 181.00
*ksuE11 182.70
*B2g36160 184.00
B2g39180 B2g38910 0.00
B2g39360 0.60
B2g39420 0.80
B2g39280 1.40
B2g36920 2.50
wmc230 4.00
swm281
B2g38440
B2g37790 5.90
B2g37770
B2g37750
B2g40930 6.10
B2g26300 7.30
gwm604 7.80
B3g19870 8.90
B2g38480 9.20
Lr75 9.60
B2g38540a
B2g39790 10.40
B2g38640
B2g37000 10.70
B2g36970
B2g36910 11.00
B2g36940
B2g36930 11.10
B2g36760
B2g36820 11.30
B2g36490 11.60
B2g36400 11.90
B2g36380 12.30
B2g36330 12.40
B2g35800 14.60
B2g39290
B2g35750
B2g35740
B2g35720
B2g35690
swm271
B2g35150 14.80
B2g35070
B2g35020
B2g34840
B2g34820
B2g34660
B2g34610
B2g34710
B2g34550 15.00
B2g34450
swm276 15.60
B2g34240 16.20
B2g33060 16.60
B2g32910
gwm11
B2g31720
gwm18 16.70
swm294
swm278
swm275
swm216 16.90
swm247 17.00
B2g28450 17.70
B2g27400
B2g27350 18.20
B2g27110
B2g26990
B2g26890
B2g26840
B2g26790
B2g26780
B2g26710
B2g26650
B2g26600 18.30
B2g30170
B2g40480
B2g26260
B2g26200
B2g26180
B2g26170
B2g26100
B2g26090
Centromere

Fig. 3.2 Comparison of the genetic map of wheat chromosome 1B of the 'Arina x Forno' RIL population and the 'Arina x Arina*LrFor*' NIL population. Marker names in pink and bold refer to the flanking markers of the introgressed region of chromosome 1B into 'Arina*LrFor*'. Markers in red were designed from the SNPs between flow-sorted Illumina sequences of 'Arina' and 'Forno' using KASPar assay. Six markers in light green belong to the introgressed region but did not map in the NIL population. Five markers in dark red were mapped in the region outside the introgressed segment in the RIL population. Four markers in dark green were designed using the wheat 90K SNP array. The marker in light blue was designed from SNPs between 'Arina' and 'Forno' using the BES of chromosome 1BS of wheat cv. 'Chinese Spring'.

3.4.2 90K SNP array as a tool for marker development

As an alternative strategy for marker development, we used the wheat 90K SNP array developed by Wang et al. (2014) for genotyping a subset of 24 lines of the BC_3F_2 NIL population along with the parents 'Arina' and 'Arina*LrFor*'. This array consists of 91,829 genome-wide gene-associated SNPs present in wheat populations of different geographical origins. Out of 91,829 SNPs, 81,587 SNPs produced functional assays (Wang et al. 2014). High density SNP array is a powerful tool to study marker-trait associations in different mapping populations and for genetic diversity analysis among wheat genotypes. The 81,587 functional SNPs were analysed for a signal in both parents i.e. 'Arina' and 'Arina*LrFor*' and 83% (67,842) of them detected a signal. Out of these 67,842 SNPs, 1.4% (965) SNPs detected a polymorphism between 'Arina' and 'Arina*LrFor*'. For the applicability of the assay on F_2 mapping populations, we expected that each SNP assay should detect 50% heterozygous alleles and show 1:2:1 segregation (1AA:2AB:1BB) in the NIL population. Chi square analysis of the 965 polymorphic SNP assays showed that 35% of the SNP assays deviate from the expected 1:2:1 segregation ratio. Nevertheless, 65% SNPs i.e. 629 SNPs detected 50% heterozygous alleles and we considered only these SNPs for further analysis.

Of the remaining 629 SNP assays (65% of 965), 226 were 1B-specific and mapped on the NIL population. Out of 226, 29 were mapped in the target region between *gwm604- swm271* (as described in chapter 2 of this thesis). For genotyping a complete set of 65 lines of the BC_3F_2 NIL population as well as the RIL population, we chose KASPar genotyping instead of the 90K SNP array because of its high cost. For KASPar genotyping we selected out of 29 SNPs, 10 SNPs that flanked the *Lr75* gene. The selected 10 SNPs were IACX5756, IACX5970, BS00003839_51, Kukri_c18052_356, Kukri_c81766_107, wsnp_Ra_rep_c74968_72704437, IAAV7575, Kukri_c17540_394, BS00049927_51 and BS00051105_51. The 100-120 bp sequence of these 10 SNPs were compared with the flow-sorted Illumina sequence assembly of cv. 'Forno'. We identified six 'Forno' contigs which had a 100% alignment with these SNP sequences (Table 3.1). After search against the *Brachypodium* protein database, these six 'Forno' contigs resulted in the identification of five proteins. The identified six 'Forno' contigs were then aligned against the flow-sorted Illumina sequence assembly of cv. 'Arina' to search for SNPs between 'Arina' and 'Forno' and to design primers flanking the SNPs for KASPar genotyping.

A set of five SNPs (IACX5970, BS00003839_51, Kukri_c18052_356, Kukri_c81766_107 and wsnp_Ra_rep_c74968_72704437) hit the same 'Forno' contig which is Taes_Forno_c-194069 (Table

3.1). This 'Forno' contig had a hit with the *Brachypodium* protein Bradi2g38480 from which a KASP marker *B2g38480* was already mapped at 0.4 cM distal to the *Lr75* gene in the NIL population (Fig. 3.2). Therefore, these SNPs were excluded from KASP genotyping. Another SNP (BS00049927_51) was also not used for KASP genotyping because 99% similarity was identified between 'Forno' and 'Arina' and the detected SNPs were present either at the beginning or at the end of the sequence in 'Forno'. Therefore, we could not use those SNPs for KASPar genotyping. In total, 4 new SNP markers, *B3g19870, B2g36970, B2g36490,* and *B2g38640* were designed which were mapped in both RIL and NIL populations using KASPar genotyping (Fig. 3.2, markers in light green colour). These markers were also named with '*B2*g' and the *Brachypodium* gene number. The introduction of these 4 new markers in the genetic map resulted in the placement of marker *B3g19870* at 0.3 cM distal to the flanking marker *B2g38480*. The marker *B2g38640* co-segregated with the proximal flanking marker *B2g38540a* at a distance of 0.8 cM proximal to the *Lr75* gene. The other two SNP markers, *B2g36970* and *B2g36490* mapped at a respective distance of 1.1 cM and 2 cM proximal to *Lr75* (Fig. 3.2).

Table 3.1 Ten SNP-based markers chosen from 90K SNP array for KASPar genotyping

SNP-based markers from 90K SNP array that flank the gene	***Brachypodium* protein hits**	**Hit with the flow-sorted Forno contigs**	**Hit with the flow-sorted Arina contigs**	**KASP marker name**
IACX5756	Bradi3g 19870	Taes_Forno_c-412683 (8.1 kb)	Taes_Arina_c_218257	*B3g19870*
IACX5970	Bradi2g 38480	Taes_Forno_c-194069 (7 kb)	Taes_Arina_c_64478	*
BS00003839_51	Bradi2g 38480	Taes_Forno_c-194069 (7 kb)	Taes_Arina_c_64478	*
Kukri_c18052_356	Bradi2g 38480	Taes_Forno_c-194069 (7 kb)	Taes_Arina_c_64478	*
Kukri_c81766_107	Bradi2g 38480	Taes_Forno_c-194069 (7 kb)	Taes_Arina_c_64478	*
wsnp_Ra_rep_c74968_72704437	Bradi2g 38480	Taes_Forno_c-194069 (7 kb)	Taes_Arina_c_64478	*
IAAV7575	Bradi2g 36970	Taes_Forno_c-270818 (6.6 kb)	Taes_Arina_c_105737	*B2g36970*
Kukri_c17540_394	Bradi2g 36490	Taes_Forno_c-149228 (2.6 kb)	Taes_Arina_c_729515	*B2g36490*
BS00049927_51	Bradi1g 36060	Taes_Forno_c-237969 (771 bp)	Taes_Arina_c_208795	No good SNP
BS00051105_51	Bradi2g 38640	Taes_Forno_c-538727 (1.15 kb)	Taes_Arina_c_794944	*B2g38640*

*KASP marker *B2g38480* was already designed from the 'Forno' contig 194069 and mapped at 0.4 cM distal to *Lr75* in the NIL genetic map. Hence, we did not include it for SNP development.

3.4.3 Use of BAC end sequences (BES) of wheat chromosome 1BS to design one new proximal flanking marker

As described above, we identified gene-associated chromosome 1B-specific SNPs by using the sequenced flow-sorted chromosomes 1B of 'Arina' and 'Forno'. In order to anchor the *Lr75* flanking SNP markers to the physical map, we used the BAC end sequences (BES) of chromosome 1BS of wheat cv. 'Chinese Spring' (Raats et al. 2013). To start with, we searched for the sequences of the four flanking markers (*B3g19870, B2g38480, B2g38640* and *B2g39790*) in the 1BS BES of wheat cv. 'Chinese Spring'. Interestingly, from the four flanking markers, sequences of the two distal flanking markers (*B3g19870* and *B2g38480*) were identified on one contig 85 (1.3 Mb) of scaffold 53 which is 15 Mb in size. The presence of five recombinants between *Lr75* and proximal flanking marker

B2g39790 should allow to add more markers proximally. We therefore decided to use the BES identified in BACs of the minimum tiling path (MTP) of this scaffold to design more markers with the goal of getting a closer proximal flanking marker from the same scaffold. To design the primers, we aligned all the BES of the BAC clones present in the MTP against the flow-sorted 'Forno' assembly using BLASTN command. The 'Forno' contigs with 98-100% sequence similarity were further aligned against the *Brachypodium* coding sequences to identify the gene-associated 'Forno' contigs. The identified 'Forno' contigs were compared with the flow-sorted 'Arina' sequence assembly to identify SNPs between the two cultivars. Using this information three primers were designed from the SNPs between 'Arina' and 'Forno' and were genotyped using high resolution melting (HRM) platform (Table 3.2). Out of these, one marker, *B2g38540a* (Fig. 3.2, marked in light blue) mapped at a distance of 0.8 cM proximal to the *Lr75* gene and co-segregated with the markers, *B2g39790* and *B2g38640*. The other two primers did not differentiate clearly between the heterozygotes and hence were not used for genotyping. By using this approach, we were not able to reduce the distance between *Lr75* and proximal flanking markers *B2g38640* and *B2g39790* but we got one additional proximal flanking marker, *B2g38540a* which is present on the same scaffold as the distal flanking marker. This helped us to establish the physical target interval as described in the next chapter (chapter 4) of this thesis.

Table 3.2 HRM primers designed from SNPs between 'Forno' and 'Arina' contigs which were identified using BES of chromosome 1BS

HRM primer name	**Forward Primer sequence 5'→3'**	**Reverse Primer sequence 5'→3'**	**Forno contig (size)**	**Arina contig (size)**	***Brachypodium* gene hit**	**SNP or InDel (Forno/Arina)**
B2g38540a	TTTAACGGAAGTCCCTCACG	ATGGCGGTGGCTACTGTACT	Taes_Forno_c_173215 (6.6 kb)	Taes_Arina_c_211461 (6.6 kb)	*Bradi2g38540*	G/A
B2g46420	TATGGGCAACTATGGCACCT	TCTGTGAACCCCTCCAAAAG	Taes_Forno_c_53259 (3.6 kb)	Taes_Arina_c_203576 (4.6 kb)	*Brdi2g46420*	A/0
B3g45660	GACCAATGTGGTTAGCTCAGTC	CCAGGTTGGTTATATGCCTCTG	Taes_Forno_c_95123 (5 kb)	Taes_Arina_c_366872 (3.1kb)	*Bradi3g45660*	T/C

In summary, by using three different strategies for marker development, i.e. flow-sorted sequenced chromosomes 1B of wheat cv. 'Arina' and 'Forno', the wheat 90K SNP array and 1BS BES of wheat cv. 'Chinese Spring', we mapped in total 78 SNP markers on the RIL population and 68 SNP markers on the NIL population. Of the 78 SNP markers, 74 (Fig, 3.2, RIL map, markers in red, light green and dark red colour) were developed from the flow-sorted sequenced chromosomes 1B of 'Arina' and 'Forno' and four SNP markers (Fig. 3.2, RIL map, markers in dark green colour) were developed from the wheat 90K SNP array. In the NIL map, out of 68 SNP markers, 63 (Fig. 3.2, NIL map, markers in red colour) were developed from the flow-sorted sequenced chromosome 1B of 'Arina' and 'Forno', four SNP markers (Fig. 3.2, NIL map, markers in dark green colour) were developed by using the wheat 90K SNP array and one SNP marker (Fig. 3.2, NIL map, marker in light blue colour) was developed from the 1BS BES of wheat cv. 'Chinese Spring'. All these strategies led to the development of genome-specific, gene-associated SNPs which resulted in the establishment of high resolution genetic map of wheat chromosome 1BS where *Lr75* gene is present. The addition of

new markers placed the *Lr75* gene between the markers *B2g38480* and *B2g38540a* within an interval of 1.2 cM.

3.4.4 Collinearity of the marker order between genetic maps of the RIL and NIL populations

To determine the collinearity of the marker order between RIL and NIL populations, the genetic linkage maps of both populations were compared (Fig. 3.2). From the 78 mapped markers in the RIL population, 73 were mapped in the region between the markers *barc128-gwm131* in the RIL population (Fig. 3.2, markers marked in red and light green). The region from *barc128-gwm131* was the region which was introgressed into 'Arina*LrFor*'. The remaining five markers were mapped outside this region in the RIL population (Fig. 3.2, markers in dark red in RIL map) and these markers were monomorphic in the NIL population. With the exception of six of the 73 mapped markers (*B2g36730*, *B2g35590*, *B2g34970*, *B2g34490*, *B2g33570* and *B2g40820*) in the region between the markers *barc128-gwm131* (Fig. 3.2, markers in light green in RIL map), the other 67 markers were mapped in the NIL population. These six markers were expected to also map in the NIL population because they are located in the introgressed region. Genotyping errors in the NIL population such as: 1) failure of amplification in 'Arina*LrFor*' (*B2g34490*), 2) failure to detect the heterozygous genotype (*B2g34970*, *B2g36730*, *B2g35590*), 3) presence of the same allele in both parents which was most likely caused by the wrong allele calling by KASPar assay (*B2g33570*, *B2g40820*) are the different reasons why these markers could not be mapped in the NIL population. As expected, perfect collinearity of the marker order was observed between the genetic linkage maps of both populations. Several groups of co-segregating markers were observed in both populations.

3.4.5 Screening of wheat germplasm with a subset of developed SNP markers to determine the origin of *Lr75*

The SNP markers developed during this study were from the small genomic region of chromosome 1BS where *Lr75* gene is located. To the best of our knowledge, the *Lr75* gene has only been identified in 'Forno' and has never been described in any other wheat cultivar. Therefore, to determine the origin of *Lr75*, we genotyped 43 wheat lines from different areas of the world with a main focus on the European wheat germplasm by using a subset of 66 SNP markers. Varieties from the European spring and winter wheat germplasm and from germplasm of the United States, Canada, Mexico and China were included. Out of the 43 wheat lines, 32 lines were of European origin, four from USA, three from Mexico, three from China and one from Canada were included. Out of the 32 wheat lines of European origin, 17 were from Switzerland, six from Germany, four from France, two from Austria, one from Sweden and two lines were from the United Kingdom (Appendix F).

The 66 SNP markers displayed an average polymorphism information content (PIC) of 0.28 ± 0.16 with PIC values ranging from 0.09 to 0.57 (Appendix G). The average heterozygosity level (0.04

± 0.1) of the markers within the 43 wheat lines was low, with 37 markers showing only the homozygous alleles and 29 markers with a heterozygosity level between 0.02 and 0.79. For the 66 analyzed markers, the major allele frequency of 'Arina' and 'Forno' alleles ranged between 0.50 and 0.98 (Appendix G). Of the 66 SNP markers, 6 SNP markers (*B2g26100, B2g27110, B2g28450, B2g39290, B2g39360* and *B2g40930*) only detected the 'Arina' allele in the tested germplasm (Appendix H). Overall, the average frequency of the 'Arina' allele was higher than the 'Forno' allele in all the tested genotypes (Appendix G, Appendix H). Also, the frequency of 'Arina' and 'Forno' alleles was calculated in 43 genotypes individually (Table 3.3, Fig. 3.3). Notably, the two Swiss winter wheat lines 'Forel' and 'Fiorina' contained a higher amount of the 'Forno' alleles (~60 %), whereas in the five lines 'Tapidor', 'Winnetou', 'Disponent', 'Transec' and 'Cambrena' a large number of markers did not amplify (~25 %) (Fig. 3.3). We can therefore conclude that the composition of chromosome 1B of the latter five lines is very different from the one of 'Arina' and 'Forno'. From this analysis, we conclude that the developed SNP markers show a wide range of polymorphism in wheat germplasm and can be used as a valuable source to detect polymorphism between different wheat cultivars as well as to determine wheat genotypes with similar haplotypes in the *Lr75* region.

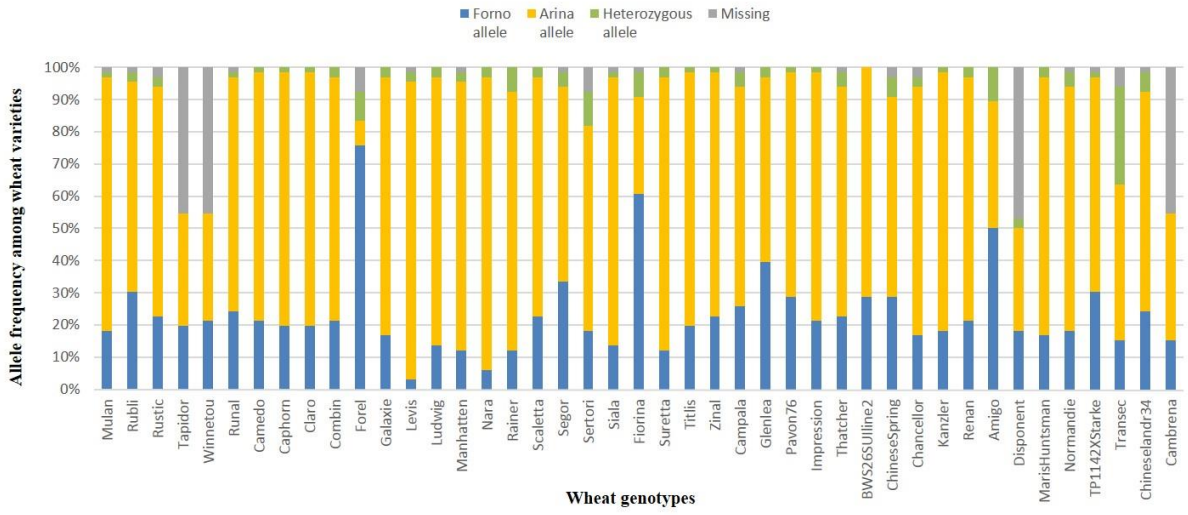


Fig. 3.3 The graphical representation of the frequency of alleles in 43 genotypes individually. The horizontal axis gives the names of the genotypes. The vertical axis shows the percentage of 'Arina', 'Forno' and heterozygous allele in each genotype. Yellow bars represent the 'Arina' allele, blue bars represent the 'Forno' allele, green bars represent the heterozygous allele and grey bars represent missing data.

Table 3.3 Frequency of the 'Arina', 'Forno' and heterozygous allele in 43 genotypes

Country	Genotypes	Forno allele	Arina allele	Heterozygous allele	Missing	Total
GER	Mulan	12	52	1	1	66
CH	Rubli	20	43	2	1	66
FRA	Rustic	15	47	2	2	66
FRA	Tapidor	13	23	0	30	66
GER	Winnetou	14	22	0	30	66
CH	Runal	16	48	1	1	66
CH	Camedo	14	51	1	0	66
UK	Caphorn	13	52	1	0	66
CH	Claro	13	52	1	0	66
CH	Combin	14	50	2	0	66
CH	Forel	50	5	6	5	66
FRA	Galaxie	11	53	2	0	66
CH	Levis	2	61	2	1	66
AUT	Ludwig	9	55	2	0	66
GER	Manhatten	8	55	2	1	66
CH	Nara	4	60	2	0	66
AUT	Rainer	8	53	5	0	66
CH	Scaletta	15	49	2	0	66
CH	Segor	22	40	3	1	66
CH	Sertori	12	42	7	5	66
CH	Siala	9	55	1	1	66
CH	Fiorina	40	20	5	1	66
CH	Suretta	8	56	2	0	66
CH	Titlis	13	52	1	0	66
CH	Zinal	15	50	1	0	66
CH	Campala	17	45	3	1	66
CAN	Glenlea	26	38	2	0	66
MEX	Pavon76	19	46	1	0	66
GER	Impression	14	51	1	0	66
MEX	Thatcher	15	47	3	1	66
MEX	BWS26SUIline2	19	47	0	0	66
CHN	ChineseSpring	19	41	4	2	66
USA	Chancellor	11	51	2	2	66
GER	Kanzler	12	53	1	0	66
FRA	Renan	14	50	2	0	66
USA	Amigo	33	26	7	0	66
GER	Disponent	12	21	2	31	66
UK	MarisHuntsman	11	53	2	0	66
USA	Normandie	12	50	3	1	66
SWE	TP1142XStarke	20	44	1	1	66
USA	Transec	10	32	20	4	66
CHN	Chineselandr34	16	45	4	1	66
CHN	Cambrena	10	26	0	30	66

3.5 Discussion

In our study we have used different strategies for marker development in order to enrich the genetic linkage maps of 'Arina x Forno' and 'Arina x Arina*LrFor*' derived RIL and NIL mapping populations, respectively, in the *Lr75* region of chromosome 1B. The different strategies used were the flow-sorted sequenced chromosomes 1B of wheat cultivars 'Arina' and 'Forno', the wheat 90K SNP array and BES of chromosome 1BS of wheat cv. 'Chinese Spring'. Using these strategies, we mapped 78 SNP markers on the RIL population and 68 SNP markers on the NIL population, respectively. The addition of 68 SNP markers reduced the *Lr75* target region from 4.3 cM (as described in chapter 2 of this thesis) to 1.2 cM with *B2g38480* and *B2g38540a* as the distal and proximal flanking markers, respectively, of *Lr75*.

3.5.1 High specificity of gene-associated SNP markers derived from flow-sorted chromosome sequences

Due to the presence of more than 80% of repetitive elements, mapping in hexaploid wheat is complicated and a challenging task. In addition, the presence of homeologous gene copies in the three sub-genomes as well as paralogous copies on the same sub-genome makes mapping even more difficult. Therefore, it becomes essential to develop sub-genome specific markers to map genes of agronomical importance in hexaploid wheat. In our study we used the flow-sorted sequences of chromosome 1B of wheat cultivars 'Arina' and 'Forno'. Flow-sorting of mitotic chromosomes of cultivars 'Arina' and 'Forno' resulted in a flow karyotype or histogram generating three composite peaks (Vrána et al. 2000). One of the peaks contains chromosome 1B mixed with eight other chromosomes (2A, 3A, 4A, 4B, 5A, 5B, 6B and 7A). Most importantly, the homeologous 1A and 1D chromosomes were not enriched in this fraction. We obtained an average purity of 25% for chromosome 1B in both cultivars. Though the purity of chromosome 1B was only 25%, the lack of homeologous chromosomes made this approach highly useful for developing genome-specific sequence-based SNP markers.

Shatalina et al. (2013) used the same approach for marker development on chromosome 3B. As chromosome 3B is the largest wheat chromosome, it forms a distinct peak in the flow karyotype and it is relatively easy to isolate pure chromosome 3B as compared to other wheat chromosomes. In their study gene-associated genome-specific SNPs were designed from the flow-sorted chromosome 3B of wheat cv. 'Arina' and 'Forno' and highly stringent criteria for SNP identification were used. In total, 70 out of 95 SNPs were polymorphic between 'Arina' and 'Forno' out of which 94% were genome-specific. Following a similar strategy, we also developed genome-specific, gene-associated SNP markers in order to increase the marker specificity of chromosome 1B. To reduce the nonspecific amplification, we masked all the repetitive regions and focussed on the SNPs which were positioned upstream or downstream of coding regions. Further, we selected only those contigs for SNP

identification which had non-redundant gene hits. Because of these stringent parameters we achieved a success rate of approximately 67% (74 out of 110) and 56% (63 of 110) for mapping SNP markers in RIL and NIL mapping populations, respectively. We consider the number of identified SNPs as a good success because of the low level of varietal SNP polymorphism in wheat as shown in previous studies (Barker and Edwards 2009). In wheat, the majority of sequencing analysis have been conducted on cv. 'Chinese Spring'. Due to the availability of ditelosomic lines of 'Chinese Spring', each chromosome arm can be isolated using flow cytometry sorting (Doležel et al. 2007). However, due to the non-availability of cytogenetic stocks, sorting of chromosome arms in other wheat cultivars is not possible. Recently, Vrána et al. (2015) have widened the applicability of chromosome flow-sorting in wheat cultivars other than 'Chinese Spring'. Using this approach, they dissected the composite chromosome peaks into smaller sections comprising of only few chromosomes. The chromosome composition in the sorted fraction was then analyzed using FISH with fluorescently labelled probes. This resulted in sorting of the sub-genomic fractions comprising of few chromosomes with one chromosome being present in higher abundance than other chromosomes in each sorted fraction. The purity of the dominant chromosome ranged from 16% (chromosome 7B) to 80% (chromosome 2B). They showed that out of 21 chromosomes 15 can be sorted without being contaminated by homeologous chromosomes. Such high purity of chromosome fractions further reduces the DNA complexity and results in the targeted development of genome-specific SNP markers with higher efficiency.

Marker development strategies using sequenced flow-sorted chromosomes can also serve as a powerful tool for diversity analysis studies by genotyping different wheat varieties or populations. In our study, we developed gene-associated SNPs from the non-coding regions of the genes. The non-coding regions are less conserved than the coding regions and this results in a higher rate of polymorphism between different wheat varieties as compared to the SNPs present in coding regions. We analyzed the diversity of a subset of 66 SNP markers from the *Lr75* region on 43 different wheat genotypes from all over the world. This analysis allowed us to determine the potential of the genome-specific SNP markers in a wide germplasm as well as polymorphism frequency in the wheat germplasm. The results obtained during this study demonstrates a wide applicability of the developed SNPs in wheat germplasm. Shatalina et al. (2013) reached the same conclusion after studying the frequency of polymorphism of a subset of 48 SNPs developed for 'Arina' and 'Forno' on the same set of 43 wheat lines used in this study. In addition, the screening of wheat lines with the SNP markers present in the *Lr75* target region gives an information about the origin of *Lr75*. We conclude that the strategy to develop gene-associated SNPs from the sequenced flow-sorted chromosomes is a useful and a highly efficient tool for high resolution genetic mapping.

3.5.2 90K SNP array as a tool for high resolution mapping

The wheat 90K SNP array represents a useful resource for high density genotyping in wheat, for studying the association between genotype and phenotype and is also useful for diversity analysis among different wheat varieties and landraces (Wang et al. 2014). In our study we used this array to genotype a subset of 24 NIL lines along with the parents, 'Arina' and 'Arina*LrFor*' with the aim of testing the array on a heterozygous population. The array consists of 81,587 functional SNPs as the remaining 8,413 SNPs for some reason never worked. Out of 81,587 SNPs, 67,842 SNPs were called in our study with the success rate of 83%. Wang et al. (2014) used the wheat 90K SNP array to study the genetic variation in allotetraploid and allohexaploid wheat populations. They used a combination of six different double haploid mapping populations and a consensus map consisting of 40,267 SNPs was constructed. In our study we used a BC_3F_2 population which consisted of 50% individuals with heterozygous alleles. Each of the mapped SNP markers in our study was analysed for the detection of heterozygous alleles in the NIL lines. The analysis revealed that out of 965 mapped SNPs, 35% SNPs detected less than 50% heterozygous alleles demonstrating that the SNP array is not optimal for heterozygous detection. SNP arrays have also been developed for highly heterozygous and diploid species such as apple (*Malus* x *domestica* Borkh). Bianco et al. (2014) developed the 20K SNP array from 13 apple cultivars and genotyped 21 full-sib families. The validation of 88% SNPs out of 18,019 SNPs demonstrated the successful application of the 20K SNP array in apple. Other than apple, SNP arrays have also been successfully used for *Eucalyptus* species which again have highly heterozygous genome (Grattapaglia et al. 2011). Therefore, heterozygosity not necessarily has to be a problem for SNP arrays. The fact that 35% of the mapped SNPs failed to detect heterozygous alleles might be due to the combination of hexaploid and heterozygous nature of wheat lines used in this study. Based on our analysis, we conclude that 90K SNP array might not be considered as a useful tool for reliably detecting heterozygous alleles.

'Arina*LrFor*' consists of 6.25% of the genome from 'Forno' including the *Lr75* target region. Therefore, it might be interesting to know which additional chromosomal segments are derived from 'Forno' by constructing the genetic map from the remaining 403 polymorphic SNP markers. Most interesting will be to know which SNPs belong to chromosome 7BL because the second resistance locus, *QLr.sfr-7BL* which interacts with *Lr75* is present on 7BL. This will be helpful to use those SNPs for mapping the second resistance locus. Information about the chromosomal location of the SNPs will help us to know if other genomic segments from 'Forno' consists of some other loci which might influence the phenotype of 'Arina*LrFor*'.

3.5.3 Advantages of different approaches used in this study for marker development

In our study we used three approaches, sequenced flow-sorted chromosomes, the 90K SNP array and BES of chromosome 1BS of wheat cv. 'Chinese Spring' to develop SNP markers for high

resolution genetic mapping. Due to the polyploid nature and presence of repetitive elements, high-resolution genetic mapping and map-based cloning is a tedious task in wheat. The difficulty lies in the development of genome-specific markers between different wheat varieties and the applicability of the markers for mapping in heterozygous populations. Flow-sorting of individual wheat chromosome or chromosome arms of wheat cultivars is a powerful tool to reduce complexity of the wheat genome. We sequenced the flow-sorted chromosome 1B with 25% purity of wheat cv. 'Arina' and 'Forno' to 16-fold coverage. This resulted in the establishment of the SNP database of complete chromosome for both cultivars. Then, the use of syntenic gene homologs of *Brachypodium* resulted in the development of targeted genome-specific SNPs. The advantage of the developed genome-specific SNPs in our study was their applicability to map the heterozygotes in the BC_3F_2 NIL mapping population. In addition, the wheat 90K SNP array is a very efficient approach for detecting homozygous alleles. In our opinion, the ideal way of using these different approaches for high resolution mapping projects is to combine them. For example, it will be more reasonable to genotype only the parental lines using 90K SNP array and identify the chromosome specific SNPs by comparing the called SNPs with the available genetic map constructed by Wang et al. (2014) using six double haploid mapping populations. The SNPs of interest will then can be genotyped on the whole set of population by using KASPar assay. The combined strategy will therefore allow to increase the resolution of the genetic maps in a cost and time efficient manner.

Chapter 4

Establishment of a physical contig spanning the genetic interval containing the gene *Lr75* by using the chromosome 1BS physical information of wheat cv. 'Chinese Spring'

4.1 Summary

Lr75 is a partial, adult plant resistance gene against leaf rust and is present on chromosome 1BS in wheat. High resolution genetic mapping using different marker development strategies located *Lr75* between markers *B2g38480* and *B2g38540a*. For physical mapping of *Lr75*, 1BS physical map information of wheat cv. 'Chinese Spring' lacking the *Lr75* gene was exploited. The flanking marker sequences were compared against the chromosome 1BS-specific BAC end sequences (BES) of the fingerprinted BAC clones obtained by both FingerPrinted Contig (FPC) and Linear Topology Contig (LTC) assembly of wheat cv. 'Chinese Spring'. The flanking marker sequences identified two BAC clones both present in scaffold 53 which had a total size of 15 Mb. In total 24 BAC clones which are all present in the minimum tiling path (MTP) of 5 BAC contigs in scaffold 53 were identified in the region between the two BAC clones containing the flanking marker sequences. Sequencing and manual assembling of the 24 BAC clones resulted in the establishment of a 3.0 Mb physical contig. The *Brachypodium* protein database was used to identify the gene-associated BAC sequences which were then compared against the flow-sorted sequence assembly of cultivars 'Forno' and 'Arina' to identify SNPs between the two cultivars. The identified SNPs were genotyped using KASPar genotyping. Of the 32 SNP markers designed, 15 markers were polymorphic between 'Arina' and 'Arina*LrFor*'. Addition of the 15 new SNP markers to the genetic map led to the establishment of a physical target region of 723 kb in 'Chinese Spring'. To establish the physical target region of *Lr75*, a non-gridded BAC library with 1X coverage from *Lr75* donor cv. 'Forno' was constructed. The assembly and annotation of the BAC sequence contigs from the 'Forno' BAC library resulted in the establishment of 403.6 kb physical target sequence with one gap. Comparison of the two haplotypes of 'Chinese Spring' and 'Forno' revealed a series of insertion, deletion and duplication events, thereby showing a considerable divergence between two haplotypes. The annotation of the physical regions of both haplotypes resulted in the identification of 14 genes in 'Chinese Spring' and eight genes in 'Forno'. In this study, we demonstrate the use of chromosome 1BS-specific BES obtained from FPC and LTC assembly of cv. 'Chinese Spring' as well as use of non-gridded BAC library of cv. 'Forno' in order to establish a physical target region spanning the *Lr75* locus.

4.2 Introduction

Bread wheat (*Triticum aestivum* L.) is one of the world's most important cereal crops with 713 million tonnes of wheat production worldwide (FAOstat 2014). Wheat is under constant attack of various viral, bacterial and fungal pathogens. Among them wheat leaf rust caused by the fungal pathogen, *Puccinia triticina* L. is a major threat worldwide. The use of resistance genes for breeding not only reduces the cost of fungicide use but is also an environment friendly method (Pink 2002). However, most of the released disease resistant varieties lost their resistance over a period of time due to the evolution of virulent strains of the pathogen. This results in the need for breeding cultivars which show long term resistance, or in other terms have "durable" resistance (Johnson 1981).

Seedling resistance genes are effective at all the stages of the plant in contrast to adult plant resistance (APR) genes which are effective only at the adult plant stage. APR genes mostly give only partial resistance in the field and they do not show a hypersensitive response. Instead the disease develops more slowly, therefore reducing the selection pressure on the pathogen to evolve. So far, only few such partial rust APR genes, such as *Lr22a, Lr34, Lr46, Lr67, Lr68, Yr36* and *Sr2* have been described in the literature (Singh et al. 1998; Kota et al. 2006; Hiebert et al. 2007; Fu et al. 2009; Krattinger et al. 2009; Herrera-Foessel et al. 2011; Herrera-Foessel et al. 2012; Moore et al. 2015).

We identified and characterized a new partial adult plant leaf rust resistance gene, *Lr75* which initially had been identified as a QTL, *QLr.sfr-1BS*, in the Swiss winter wheat cv. 'Forno' (Schnurbusch et al. 2004). The locus explained 28-32% of the phenotypic variation and is present on chromosome 1BS. For genetic mapping of *Lr75*, we generated a near-isogenic line, 'Arina*LrFor*' and back-crossed it to the susceptible cv. 'Arina' to construct a fine mapping population of a cross of 'Arina x Arina*LrFor*' (as described in chapter 2 of this thesis). Different strategies for marker development were used (as presented in chapter 3 of this thesis) to map *Lr75* genetically between the markers *B2g38480* and *B2g38540a* at a distance of 0.4 and 0.8 cM distally and proximally, respectively.

High resolution genetic and physical mapping of wheat has been difficult due to its large genome size (approximately 17 Gb). Hexaploid wheat (2n=6x=42) originates from two hybridization events. The first hybridization occurred approximately 0.5 million years ago between diploid relatives *Triticum urartu* (AA) and an extinct relative of *Triticum speltoides* (BB) which resulted in allotetraploid wheat, *Triticum turgidum* (AABB). The hybridization of *T. turgidum* with the D genome donor of hexaploid wheat, *Aegilops tauschii* occurred nearly 9,000 years ago and produced hexaploid bread wheat (*T. aestivum*, AABBDD). The genome of hexaploid wheat is three times the size of the barley genome and is more complex due to the presence of homeologous genes each descended from the ancestral diploid species. Furthermore, the wheat genome consists of more than 80% of repetitive elements with the presence of a very small amount of coding sequences.

With the aim to physically map and sequence the complex genome of bread wheat, the International Wheat Genome Sequencing consortium (IWGSC; www.wheatgenome.org) used the chromosome-by-chromosome approach to isolate individual chromosomes and chromosome arms by using flow cytometry sorting (Šafář et al. 2004; Safár et al. 2010). Recently, sequencing of individual chromosome arms led to the construction of a draft genome sequence of hexaploid bread wheat cv. 'Chinese Spring' (The International Wheat Genome Sequencing Consortium 2014). The genome consists of 124,201 gene loci which are distributed evenly throughout the subgenomes and homeologous chromosomes.

The development of chromosome-specific BAC libraries, the use of diploid relatives as models for sequence comparisons among the genomes as well as exome sequencing facilitated the identification of polymorphisms as well as the development of high-throughput sequence based markers. Nowadays, BAC libraries are available for all chromosomes and chromosome arms of wheat cv. 'Chinese Spring' (Molnár et al. 2014). Using these sources, a physical map of chromosome 1BS of wheat cv. 'Chinese Spring' was generated by Raats et al. (2013). The 1BS physical map consists of a LTC assembly of 32,231 fingerprinted BAC clones into 1,057 contigs. The contigs were merged to produce 57 long scaffolds with an average size of 4.6 Mb, covering 83% of the short arm of chromosome 1B. The advances in high-throughput genotyping platforms such as KASPar (Kompetitive Allele Specific PCR) and High Resolution Melting (HRM) have facilitated the mapping of sequence-based markers by genotyping single nucleotide polymorphisms (SNPs). Recently, by using EST sequences and chromosome survey sequences, several thousand SNPs have been successfully genotyped (Akhunov et al. 2010; Allen et al. 2011). All these advancements facilitate map-based cloning of disease resistance genes as well as other genes of agronomical importance in crops such as wheat. An additional approach to generate genome-specific markers was published by Shatalina et al. (2013). There, the flow-sorted chromosome 3B of wheat in combination with KASPar genotyping was used to generate a SNP map consisting of 70 KASP-based SNP markers in the 'Arina x Forno' recombinant inbred line (RIL) population.

Here, we report on the establishment of a physical target interval spanning the *Lr75* locus by using the chromosome 1BS-specific BES sequences of the fingerprinted BAC clones obtained by both FPC and LTC assembly of wheat cv. 'Chinese Spring'. A physical region of 3 Mb which consisted of 24 BAC clones was established for 'Chinese Spring'. To map markers derived from these 24 BAC clones on the NIL genetic map, SNPs between the flow-sorted sequences of 'Forno' and 'Arina' were identified and genotyped using KASPar genotyping. Mapping of the SNPs resulted in the establishment of a 723 kb physical sequence contig from cv. 'Chinese Spring'. In addition, we also established a physical sequence contig of 403.6 kb from the resistant cv. 'Forno' by using a 'Forno' non-gridded BAC library with 1X coverage. Annotation of the physical sequence contigs of both

'Chinese Spring' and 'Forno' based on the *Brachypodium* protein database led to the identification of 14 and 8 candidate genes in 'Chinese Spring' and 'Forno', respectively.

4.3 Material and Methods

4.3.1 Plant material

Two Swiss winter wheat cultivars, 'Arina' and 'Forno' were used to generate the SNP-based KASP markers in this study. DNA extraction was done as per the protocol described by Stein et al. (2001). For physical mapping of PCR-based marker as well as for screening BAC libraries, nullitetrasomic line of chromosome 1B, N1BT1D was used to test for B-genome specificity.

4.3.2 Screening of the cv. 'Chinese Spring' BAC library and BAC DNA isolation

For physical mapping of the *Lr75* locus, the *Triticum aestivum* cv. 'Chinese Spring' 1BS specific BAC library (TaaCsp1BShA) generated by the French Plant Genomic Resource Centre (CNRGV), France was used (http://cnrgv.toulouse.inra.fr/).

BAC clone cultures of the TaaCsp1BShA BAC library received from CNRGV, France were streaked on solid LB medium supplemented with chloramphenicol (12.5 mg/ml) and kept at 37°C overnight. A single colony per BAC clone was picked and cultured in a 5 ml culture and kept overnight in a shaking incubator at 37°C. The next day, a 250 ml culture was prepared by adding 250 µl of the starting culture and also kept overnight in a shaking incubator at 37°C. BAC DNA was extracted using the QIAGEN® Large-Construct Kit (Qiagen, Venlo, Netherlands).

4.3.3 BAC clone sequencing and sequence assembly of cv. 'Chinese Spring'

A minimum tiling path (MTP) consisting of 24 BAC clones spanning the *Lr75* region of cv. 'Chinese Spring' was sequenced by Microsynth AG, Balgach, Switzerland using MiSeq paired-end sequencing in 2*250 cycles. For *de novo* sequence assembly of BAC clone reads into contigs, CLC Main Workbench 6.5.2 with standard settings was used. Only contigs with coverage above 500x were selected for construction of the physical map.

4.3.4 Orientation and positioning of assembled sequence contigs

Sequences from the BAC clones lying in the genetically defined target interval were manually assembled into one continuous sequence to establish a complete sequence contig spanning the region of interest. For this purpose, the sequence contigs obtained from the CLC assembly of the BAC clones were first analyzed for overlaps with sequence contigs of other BAC clones using the commands

'BLASTN' and 'dot2'. BLASTN determines the alignment between the contigs of two BAC clones. The best possible hit was then used to create a dot plot between the two sequences to analyse the exact start and end position of the overlap. The neighbouring BAC clones and their order were deduced and contigs were integrated and oriented in the physical map. To avoid false positives due to the vector sequence, the vector sequence was queried against all the contigs of each BAC clone using BLASTN search against the vector database. To assemble the contigs which did not overlap with the neighbouring BAC clones, analysis of transposable elements (TEs) based on the *Triticeae* database (TREP) (Wicker et al. 2002) was used. This strategy takes advantage of the high abundance of TEs in the wheat genome and the fact that TEs create a specific footprint, called target duplication site (TDS), upon integration into the genome. The TDS leads to identical sequence of around five nucleotides flanking the TE. BLASTN search against the TREP database (Wicker et al. 2002) was conducted for non-overlapping BAC sequence contigs of the same BAC clone. Hits of two contigs with the same TE confirmed the position and orientation of the two sequence contigs relative to each other. Exact overlap was determined using the command 'dot2' and the gap between the contigs was covered with the TE sequence. Since the same TE can be present in multiple copies, the TDS served as a confirmation that start and end belong to the same TE. BAC sequence contigs which had neither a hit with another sequence contig nor with a TE could not be oriented, and hence were integrated into the physical map at a likely position in forward orientation (Fig 4.1)**.**

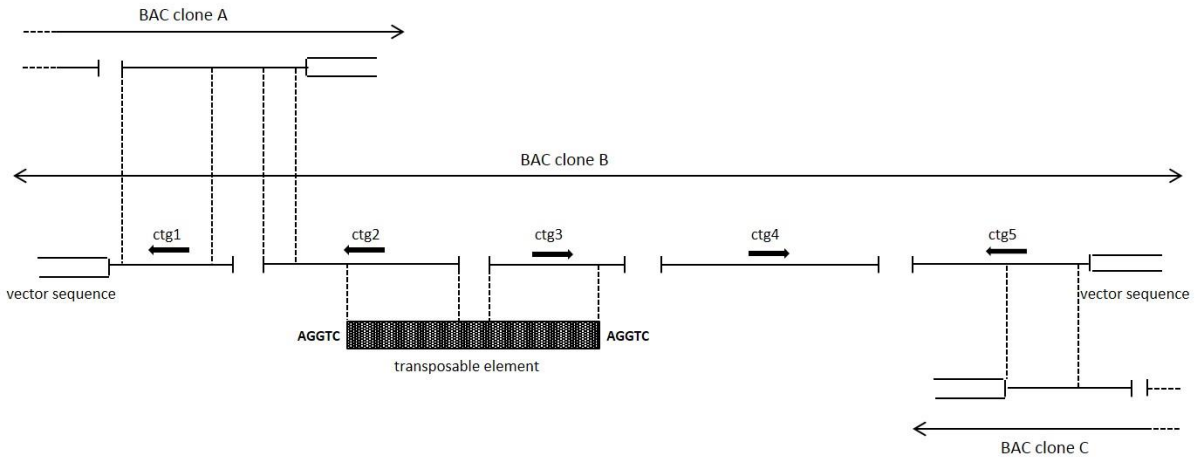


Fig. 4.1 Strategy to establish position and orientation of BAC-derived sequence contigs. The vector sequences in ctg1 and ctg5 of BAC clone B indicate the start and end of the inserts of BAC clones A, B and C. The orientation of ctg1 and ctg2 can be deduced from their overlaps with contigs of BAC clones A and C. Ctg2 and ctg3 do not overlap, but both align to the same specified transposable element (TE). Start and end position of TEs can be identified by the target duplication site (TDS). Since the same TE can be present in multiple copies, this footprint serves as a confirmation that start and end belong to the same TE. Ctg4 has neither a hit with another contig nor with a TE and its true orientation remains unclear, thus it is integrated at a likely position in forward orientation (→).

4.3.5 Development of amplification products for screening the 'Forno' BAC library

Chromosome 1B-specific probes were developed from the genic regions present in the 723 kb physical sequence of cv. 'Chinese Spring'. The gene sequences were compared against the flow-sorted

'Forno' sequence assembly using BLASTN command. PCR probes were then designed flanking the SNPs between the 'Chinese Spring' sequence and the 'Forno' sequence. The primers were tested for 1BS specificity by non-amplification of DNA from a nulli-tetrasomic line (N1BT1D) of 'Chinese Spring'. The annealing temperature was tested for each marker using gradient PCR. All PCR products were analyzed on agarose gels. For amplification the following standard protocol was used.

PCR standard protocol

13.5 µl	H_20	95°C for 4 min	
2 µl	10x PCR buffer (Sigma, Switzerland)	95°C for 30 s	
1 µl	dNTPs 2.5 mM	T_A for 30 s	35 cycles
0.5 µl	Forward primer 10 µM	72°C for 30 s	
0.5 µl	Reverse primer 10 µM		
2 µl	Template DNA (65 ng/µl)		
0.5 µl	Taq DNA polymerase 5 U/µl (Sigma, Switzerland)		
20 µl	Total reaction volume		

4.3.6 Construction of non-gridded BAC library from cv. 'Forno'

A non-gridded BAC library with 1X coverage, Tae-B-Forno-ng from cv. 'Forno' was constructed by the French Plant Genomic Resource Centre (CNRGV), France by using a protocol described in Mago et al. (2014). The library generated 250,200 BAC clones with an average insert size of 93 kb, representing a total of nearly 1.4-fold genome coverage. BAC clones were divided into 192 pools before overnight growth and DNA amplification. Screening of the pools was done by using the chromosome 1B-specific probes developed from the genic regions present in the physical region of cv. 'Chinese Spring'. The primer sequences for the probes used for 'Forno' BAC library screening are given in Appendix I.

4.3.7 BAC sequencing and assembly

The BAC clones isolated from the non-gridded 'Forno' BAC library were sequenced by CNRGV, France by using two approaches. First, a 454 GS-Junior sequencer (Titanium kit, pair-end reads) was used. In order to eliminate low quality, highly repeated reads, the raw data were first

cleaned with Pyrocleaner software and then cleaned from *E. coli* DNA sequence contamination. The remaining reads were assembled using Newbler 2.9, Roche platform. The reads having 100% match with the vector sequences were deleted before assembling. BAC clones 12E6 and 3C21 were sequenced using this approach. For the second set of sequencing, which was used to cover the gap between the two BAC clones, 12E6 and 3C21, PACBio sequencing platform was used to sequence the additional BAC clones (96M10, 123B20, 63O11 and 21I24) obtained from the 'Forno' BAC library.

4.3.8 Gene annotation

The genes were annotated in the 723 kb sequence-assembled physical region of 'Chinese Spring' as well as in the 403.6 kb sequence-assembled physical region of 'Forno' by comparing them with the *Brachypodium* CDS database using "blastf" command. The alignments with more than 85% sequence similarity were considered. The aligned CDS sequence of *Brachypodium* was then plotted against the physical model using DOTTER version 3.1 (Sonnhammer and Durbin 1996) to visualize the sequence alignments for annotating exons and to identify the exon/intron conserved sites, start and stop codons.

4.4 Results

4.4.1 Identification of a physical region spanning the *Lr75* locus using the 1BS physical map of wheat cv. 'Chinese Spring'

Lr75 was mapped on the short arm of wheat chromosome 1B flanked by the markers *B2g38480* and *B2g38540a* as described in chapter 3 of this thesis. The target region spanned 1.2 cM between the flanking markers. To establish the physical contig spanning this target interval, sequences of the flanking markers were identified in the 1BS physical map of wheat cv. 'Chinese Spring' generated by Raats et al. (2013). The 1BS physical map is comprised of 57 long scaffolds or super-contigs containing 32,231 BAC clones. The ends of the 6,447 BAC clones present in the minimum tiling path (MTP) of chromosome 1BS were sequenced by Raats et al. (2013). Sequences of the 'Forno' Illumina contigs from where the flanking markers (*B2g38480* and *B2g38540a*) were known and compared with the BAC end sequences (BES) of the 1BS contigs. The distal flanking marker sequence, *B2g38480* identified BAC clone 93G13 of BAC contig 85 which is ~1.3 Mb in size, whereas the proximal flanking marker sequence, *B2g38540a* identified BAC clone 68G23 of contig 677 which is ~1.8 Mb in size. The physical region defined by BAC contig 85 and BAC contig 677 has an additional three BAC contigs in between (ctg638, ctg542, ctg204), defining an estimated total physical target region of ~3 Mb which corresponds to 1.2 cM. All these BAC contigs fall into one long scaffold of 15 Mb in size which is named scaffold 53 (Fig. 4.2a). However, as the BAC contigs within this scaffold were described to be non-overlapping, it was not clear how big the gaps were between them.

4.4.2 Sequencing of MTP BAC clones in the 3.0 Mb target region

The minimum tiling path (MTP) potentially covering the 3 Mb target region of the *Lr75* locus was comprised of 24 BAC clones assigned to the five BAC contigs ctg677, ctg204, ctg542, ctg638 and ctg85 (Fig. 4.2a). The physical localization of these 24 BAC clones were based on the FPC assembly. The BAC clone information obtained through FPC assembly introduced some problems: 1) We observed the wrong placement of some BAC clones in BAC ctg204. In BAC ctg204, five BAC clones (97D3, 35O2, 102L8, 128K18 and 46K17) did not show any overlap with any BAC clone in the neighbouring ctg542 (Fig. 4.2a, big blue box). Instead, we found that BAC clone 23A16 of BAC ctg204 directly overlapped with BAC clone 11M16 of BAC ctg542. This indicates the wrong placement of the five BAC clones in ctg204. 2) We also observed the wrong orientation of the BAC ctg638. The overlap of BAC clone 110J10 in BAC ctg638 with the BAC clone 69B3 of BAC ctg85 indicated that the BAC ctg638 should actually be in the inverted orientation. 3) We identified several gaps while assembling the BAC clone sequences using the FPC assembly. No overlaps were observed between any BAC clone of BAC ctg542 and the BAC clones of BAC ctg638; between BAC clones 14P19 and 5L13 of BAC ctg542; between 139K17 and 93B22 of BAC ctg638 and between BAC 97L13 and 93G13 as well as 134M16 and 110D6 of BAC ctg85 (small green boxes, Fig. 4.2a).

In order to address all these problems which occurred due to the FPC assembly we explored another contig assembly program, an LTC assembly of the 1BS physical map (in collaboration with Drs. Abraham Korol, Zeev Frenkel and Tzion Fahima, Haifa University, Israel). LTC is a new program for contig assembly which resulted in the establishment of longer physical BAC contigs than those generated by the FPC program (Raats et al. 2013). LTC also facilitated the identification of poorly fingerprinted or repetitive questionable clones (Q-clones) by making a connection of significantly overlapping clones. The use of LTC assembly resulted in the identification of new BAC clones potentially covering the gaps described above (Fig. 4.2b). The additional BAC clones which were sequenced to cover the gaps were 5L12, 19B9 and 138P8 from BAC ctg542; 112H14 from BAC ctg638; 2H20 and 125P23 from BAC ctg85 (BAC clones marked in green in Fig. 4.2b). After the introduction of the new BAC clone sequences, the sequence of BAC clone 5L13 in BAC ctg542 did not show any overlap with its neighbouring clones, suggesting the wrong position of this BAC clone and it was therefore eliminated from further analysis (Fig. 4.2a, small blue box). The addition of new BAC clone sequences resulted in a final number of 24 BAC clones to get a complete physical target interval from cv. 'Chinese Spring' (Fig. 4.2b). The sequencing of these 24 BAC clones resulted in a total of 194 BAC sequence contigs. The number of sequence reads per BAC clone ranged from 357,000 (112H14) to 2,922,000 (139K17) with an average coverage of 2,361x per BAC clone. Assembly of the reads with an average length of 247 bases resulted in variable number of sequence contigs per BAC clone from one (23A16) to 24 contigs (2H20). In total, sequences having a total size of 3,203 kb were assembled (Table 4.1).

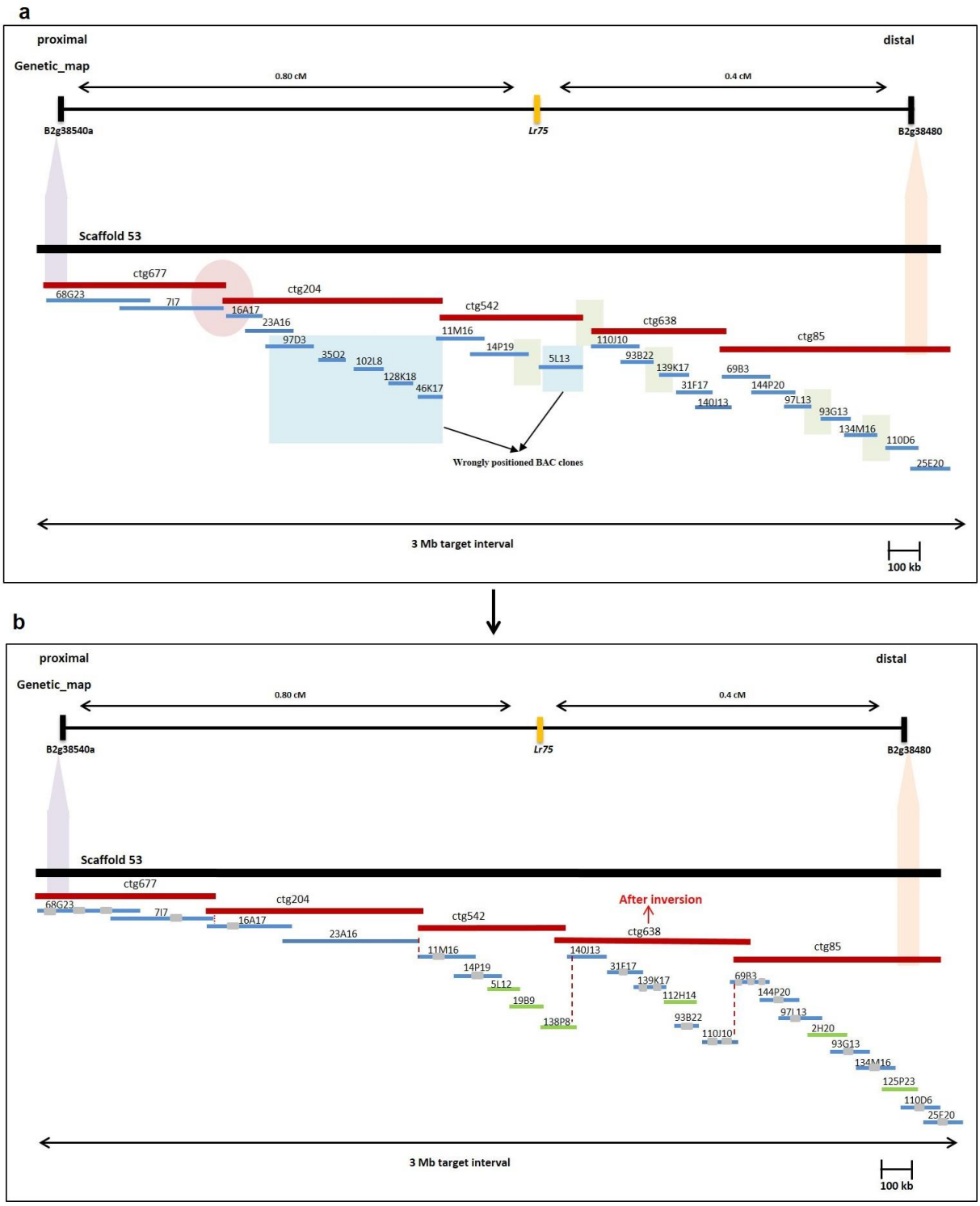

a
proximal
distal
Genetic_map
0.80 cM
0.4 cM
B2g38540a
Lr75
B2g38480
Scaffold 53
ctg677
68G23
7I7
ctg204
16A17
23A16
97D3
35O2
102L8
128K18
46K17
ctg542
11M16
14P19
5L13
ctg638
110J10
93B22
139K17
31F17
140J13
ctg85
69B3
144P20
97L13
93G13
134M16
110D6
25E20
Wrongly positioned BAC clones
3 Mb target interval
100 kb
b
proximal
distal
Genetic_map
0.80 cM
0.4 cM
B2g38540a
Lr75
B2g38480
Scaffold 53
ctg677
68G23
7I7
ctg204
16A17
23A16
ctg542
11M16
14P19
5L12
19B9
138P8
After inversion
ctg638
140J13
31F17
139K17
112H14
93B22
110J10
ctg85
69B3
144P20
97L13
2H20
93G13
134M16
125P23
110D6
25E20
3 Mb target interval
100 kb
1BS BAC contigs merged to form scaffold 53
Overlapping BAC clones in 1BS MTP BAC contig as defined by FPC assembly
New Overlapping BAC clones in 1BS MTP BAC contig as defined by LTC assembly
Genes on BAC clones

Fig. 4.2 A 3.0 Mb physical region of wheat cv. 'Chinese Spring' spanning the *Lr75* locus. Panel 'a' consists of the BAC clone sequences from FPC assembly with gaps marked with light green boxes and wrongly positioned BAC clones marked with light blue boxes. Red circle shows an overlap of BAC clone 7I7 and 16A17 as determined by sequencing, but not by FPC assembly. Blue boxes show the wrongly positioned BAC clones. Panel 'b' includes the additional BAC clone sequences obtained from the LTC assembly which are shown as green bars. These new BAC clones cover the gaps mentioned in panel 'a' and hence, established a complete physical target interval between the markers *B2g38480* and *B2g38540a*. Dark red dotted vertical lines in panel 'b' indicate the overlap between BAC clone sequences of neighbouring BAC contigs. Grey boxes on the BAC clones represent the genic sequences identified after comparing the BAC sequences with the *Brachypodium* protein database. These genic sequences were then further used to design SNP markers which were genotyped by KASPar assays.

Table 4.1 Sequence information of BAC clones in the *Lr75* target region of cv. 'Chinese Spring'

BAC contig	**BAC clone in BAC ctg**	**Estimated size** (Raats et al. 2013)	**sequence coverage x-fold**	**No. of reads**	**No. of sequence ctgs**	**Total size of sequence assembly**
ctg677	68G23	109 kb	2397	1043 k	13	113 kb
ctg677	7I7	84 kb	4353	2333 k	5	106 kb
ctg204	16A17	98 kb	1105	656 k	7	126 kb
ctg204	23A16	172 kb	1826	928 k	1	114 kb
ctg542	11M16	101 kb	1778	923 k	2	114 kb
ctg542	14P19	115 kb	892	542 k	5	120 kb
ctg542	5L12	133 kb	988	485 k	6	118 kb
ctg542	19B9	96 kb	899	463 k	3	123 kb
ctg542	138P8	89 kb	2590	1839 k	10	176 kb
ctg638	140J13	162 kb	1450	1165 k	10	168 kb
ctg638	31F17	102 kb	1602	1016 k	9	125 kb
ctg638	139K17	126 kb	4624	2922 k	6	163 kb
ctg638	112H14	127 kb	565	357 k	11	142 kb
ctg638	93B22	130 kb	841	494 k	15	134 kb
ctg638	110J10	107 kb	4357	1124 k	10	134 kb
ctg85	69B3	99 kb	2545	892 k	9	88 kb
ctg85	144P20	208 kb	2898	2479 k	11	188 kb
ctg85	97L13	160 kb	1141	810 k	4	155 kb
ctg85	2H20	113 kb	1014	1551 k	24	378 kb
ctg85	93G13	153 kb	973	1057 k	2	135 kb
ctg85	134M16	168 kb	1463	1645 k	15	155 kb
ctg85	125P23	119 kb	542	363 k	6	133 kb
ctg85	110D6	124 kb	1654	1219 k	7	143 kb
ctg85	25E20	91 kb	14166	1039 k	3	114 kb

4.4.3 SNP development from the flow-sorted Illumina sequence contigs of 'Arina' and 'Forno' to map the 24 BAC clones on the genetic map

The assembled sequence of the 24 BAC clones represents the physical region between the *Lr75* flanking markers. Therefore, these BAC clones represent an ideal source to develop additional markers in the target region for high resolution genetic mapping. In order to map sequences from these BAC clones on the genetic map of the NIL population, SNPs in genic regions were identified between the flow-sorted Illumina survey sequences of the parents 'Arina' and 'Forno'. For SNP detection, the assembled BAC sequence contigs of the 24 BAC clones were aligned against the protein database of *Brachypodium* by using the "massblastf" and BLASTX algorithm. The corresponding BAC sequence contig fragments that had more than 98% homology with the *Brachypodium* proteins were then extracted using the "assemble" command on Linux. The assembled BAC sequence contig fragment was aligned against the flow-sorted 'Forno' Illumina sequence assembly using BLASTN algorithm. The generation of the flow-sorting assemblies of cv. 'Arina' and 'Forno' was described in detail in chapter 3 of this thesis. The 'Forno' contigs having 98% or more sequence identity with the BAC sequence contig fragment were aligned with the flow-sorted Illumina sequence assembly of 'Arina'. Only alignments with 99% sequence identity between 'Arina' and 'Forno' were selected as candidate markers. SNP primer design and KASPar genotyping for the NIL population were conducted by LGC Genomics Ltd., Herts, United Kingdom. The primers were named based on the protein hit of *Brachypodium* protein database: for example, the *B2g38540b* primers were designed from the BAC sequence contig which had homology with the *Brachypodium* protein B38540 of chromosome 2. The primer sequences are given in Appendix J.

Out of 24 BAC clone nucleotide sequences, two BAC clones 23A16 (1 sequence contig) and 140J13 (10 sequence contigs) did not show any strong hit with the *Brachypodium* protein database. Hence, these two BAC clones were not used for marker development. Out of the remaining 183 sequence contigs from 22 BAC clones, 43 sequence contigs had a strong BLASTX hit with *Brachypodium* proteins. Because some protein sequences were present in multiple copies, a total of 79 different proteins were identified on the 43 sequence contigs. The particular region of the BAC sequence contig which had homology with the *Brachypodium* protein sequence was extracted and aligned against the flow-sorted 'Forno' sequence assembly. The extracted BAC sequence contig fragments had hits with 115 contigs of the flow-sorted 'Forno' Illumina sequence assembly corresponding to a total sequence size of ~122 kb. The identified 'Forno' contigs were then compared with the 'Arina' sequences for SNP detection. Not all 'Forno' contigs had hits with 'Arina' sequences. Finally, the alignment of 64 'Forno' contigs with the corresponding flow-sorted 'Arina' sequences allowed the identification of 185 SNPs and 7 insertion deletions (InDels) between the two genotypes (Table 4.2).

For KASPar genotyping, at least one SNP per protein coding sequence was selected. For historical reasons, the three BAC clones present in 1BS BAC ctg204 which were wrongly positioned

as shown in Fig. 4.2a (97D3, 35O2, 46K17) as well as BAC clone 5L13 present in BAC ctg542 were also used for primer design. Two SNPs per protein were designed for the proteins that were present in multiple copies, such as DNA/RNA helicase, Cytochrome P450, Actin binding FH2 and DRF, Signal Recognition Particle, Cyclin-like F-Box and LRR2. Finally, a total of 31 SNP-based primers were designed for 27 proteins present on 18 different BAC clones. Out of 31 SNP-based primers, 28 were amplified on both 'Arina' and 'Forno', with the success rate 90%; 15 of the amplified SNP-based primers showed polymorphism between 'Arina' and 'Forno', whereas 13 were monomorphic. Out of 13 monomorphic primers, six (*B2g36430, B2g36550, B1g600960, B2g44330a, B2g44330b* and *B4g08060*) were designed from the wrongly positioned BAC clones outside the target interval. The non-polymorphism of these six SNP primers therefore confirms the wrong position of these BAC clones. The three SNP-based primers, *B3g55220, B3g19870b* and *B1g48050* did not amplify in 'Arina' and 'Forno' and were therefore excluded (Table 4.2).

Table 4.2 Development of SNP markers from the SNPs between 'Arina' and 'Forno' for genotyping based on the KASPar assay.

1BS ctg	BAC clone	*Brachypodium protein*	Forno sequence ctg	Size (bp)	Arina sequence ctg	No. of SNPs	No. of InDels	KASP marker	Poly/ mono
Ctg677	68G23	Lipase GDSL (*B2g38570*)	Taes_Forno_c-283803	391	Taes_Arina_c-416962	no			
			Taes_Forno_c-1109320	461	Taes_Arina_c-416962	1		*B2g38570*	poly
			Taes_Forno_c-1355049	259	Taes_Arina_c-416962	no			
			Taes_Forno_c-1365629	406	Taes_Arina_c-416962		1		
		DUF246 (*B2g38540*)	Taes_Forno_c-1310777	638	Taes_Arina_c-211461	1		*B2g38540b*	poly
			Taes_Forno_c-173215	6596	Taes_Arina_c-211461	6			
					Taes_Arina_c-721018	1			
		Myb-type HTH (*B2g38560*); Purine-kinase; LRR	Taes_Forno_c-665140	808	Taes_Arina_c-717811	1		*B2g38560*	poly
			Taes_Forno_c-605864	393	Taes_Arina_c-64821	no			
			aes_Forno_c-1170056	529	Taes_Arina_c-64821	no			
			Taes_Forno_c-1029153	821	Taes_Arina_c-64821		1		
			Taes_Forno_c-1256996	372	Taes_Arina_c-1025166	no			
					Taes_Arina_c-64821		1		
			Taes_Forno_c-390952	655	Taes_Arina_c-346245	1			
	7I7	Myb-type HTH; Peptidase S8 and S 53	Taes_Forno_c-665140	808	Taes_Arina_c-717811	1			
		Peptidase A22; kelch related transcription regulator (*B2g36430*)	Taes_Forno_c-301067	238	Taes_Arina_c-789240	4		*B2g36430*	mono
Ctg204	16A17	DNA/RNA helicase (*B1g21030*)	Taes_Forno_c-112376	208	Taes_Arina_c-32174	1		*B1g21030*	mono
		Ser/Thr phosphatase	Taes_Forno_c-850048	951	Taes_Arina_c-198346	no			
	23A16	No gene							
	97D3	LRR2; Cyclin like F-box; bHLH (*B2g36550*)	Taes_Forno_c-600070	287	Taes_Arina_c-404070	1			
					Taes_Arina_c-381293	1			
			Taes_Forno_c-622693	248	Taes_Arina_c-320486	2			
			Taes_Forno_c-884783	268	Taes_Arina_c-224199	1			
					Taes_Arina_c-394029	1			
			Taes_Forno_c-83401	311	Taes_Arina_c-148914		1		
			Taes_Forno_c-640633	218	Taes_Arina_c-606854	1		*B2g36550*	mono
			Taes_Forno_c-573265	240		no			
			Taes_Forno_c-539394	260	Taes_Arina_c-691420	2			
	35O2	LRR2	Taes_Forno_c-529741	586	Taes_Arina_c-173263	2			
					Taes_Arina_c-35760	1			
					Taes_Arina_c-146886	1			
			Taes_Forno_c-322836	217	Taes_Arina_c-375328	1			
			Taes_Forno_c-454920	209	Taes_Arina_c-149475	no			
			Taes_Forno_c-87452	323	Taes_Arina_c-404070	1			
					Taes_Arina_c-34572	2			
		general ascorbate transporter B12D (*B3g46620*)	Taes_Forno_c-8469	3351	Taes_Arina_c-13163	4		*B3g46620*	mono
			Taes_Forno_c-701440	14496	Taes_Arina_c-415737	no			
			Taes_Forno_c-664766	2141	Taes_Arina_c-307720	no			
	102L8	FAE1/Type3 polyketide synthase like prot	Taes_Forno_c-670809	899	Taes_Arina_c-505181	no			
		Pentatricopeptide repeat; Peptidase S8 and S 53	No strong hit						
		Fumarate reductase; rossmann like alpha/beta/alpha sandwich	Taes_Forno_c-823497	518	Taes_Arina_c-501977	no			
		Gibberellin regulated prot; UDP-glucuronosyl low density lipo prot receptor	Taes_Forno_c-952089	820	Taes_Arina_c-973235	1			
			Taes_Forno_c-1411346	320	Taes_Arina_c-179255	no			
			Taes_Forno_c-1476409	268	Taes_Arina_c-1118083	3			
	128K18	Zn Finger/RING/FYVE/PhD type	Taes_Forno_c-418098	284	Taes_Arina_c-222383	2			
			Taes_Forno_c-289893	269	Taes_Arina_c-557884	2			
					Taes_Arina_c-585548	1			
			Taes_Forno_c-293021	1653	Taes_Arina_c-1252085	1			
					Taes_Arina_c-48090	2			
			Taes_Forno_c-26413	370	Taes_Arina_c-16749	2			
		Pentatricopeptide repeat; short chain dehydrogenase reductase	No strong hit						
		Type3 polyketide synthase	No strong hit						
		Gibberellin regulated prot; UDP-glucuronosyl low density lipo prot receptor	No strong hit						
	46K17	WD40/YVTN repeat like; aldehyde dehydrogenase (*B1g60960*)	Taes_Forno_c-129017	1821	Taes_Arina_c-55180	1		*B1g60960*	mono

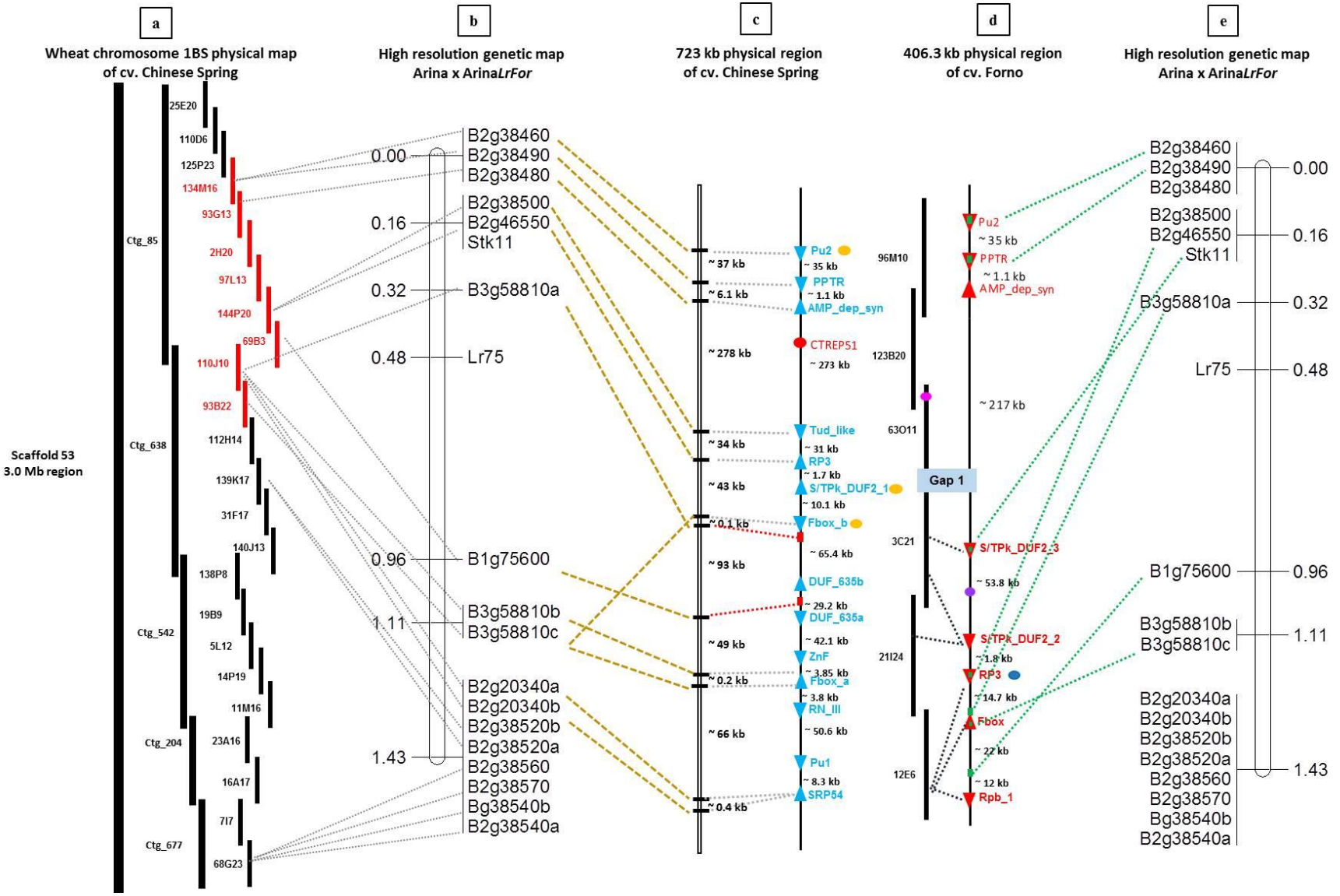
Ctg542	11M16	Alcohol dehydrogenase class 3 (*B3g55220*)	Taes_Forno_c-1377308	426	Taes_Arina_c-1278159	1		*B3g55220*	NA
		Translation elongation initiation factor	No strong hit						
	14P19	Alcohol dehydrogenase class 3	Taes_Forno_c-212743	320	No strong hit				
			Taes_Forno_c-728460	719	Taes_Arina_c-1019240	no			
			Taes_Forno_c-429165	2031	Taes_Arina_c-65266	no			
			Taes_Forno_c-980273	320	Taes_Arina_c-1215002	no			
		DNA/RNA helicase (*B1g20130*)	Taes_Forno_c-777745	301	Taes_Arina_c-215995	1			
					Taes_Arina_c-425571	1		*B1g20130b*	mono
			Taes_Forno_c-620329	229	Taes_Arina_c-215995	no			
			Taes_Forno_c-486683	211	Taes_Arina_c-522061	no			
					Taes_Arina_c-409488	1		*B1g20130a*	mono
	5L13	Cytochrome P450 (*B2g44330*)	Taes_Forno_c-1212947	258	Taes_Arina_c-80852	3			
			Taes_Forno_c-1078187	279	Taes_Arina_c-80852	no			
			Taes_Forno_c-1170573	351	No strong hit				
			Taes_Forno_c-1096100	285	Taes_Arina_c-80852	no			
			Taes_Forno_c-1167553	287	Taes_Arina_c-80852	3		*B2g44330a*	mono
			Taes_Forno_c-1212947	258	Taes_Arina_c-80852	3		*B2g44330b*	mono
			Taes_Forno_c-839279	251	Taes_Arina_c-80852	no			
		NnrU (*B4g08060*)	Taes_Forno_c-756537	3580	Taes_Arina_c-271431	no			
			Taes_Forno_c-903293	326	Taes_Arina_c-271431	1			
			Taes_Forno_c-98543	1709	Taes_Arina_c-271431	No			
					Taes_Arina_c-1242826	2		*B4g08060*	mono
					Taes_Arina_c-1189208	1			
			Taes_Forno_c-1112715	256	Taes_Arina_c-868081	no			
			Taes_Forno_c-1439100	369	Taes_Arina_c-868081	No			
Ctg638	110J10	Signal recognition particle (*B2g20340*)	Taes_Forno_c-1098139	791	Taes_Arina_c-137542	2	1		
			Taes_Forno_c-1061633	339	Taes_Arina_c-137542	1	1		
			Taes_Forno_c-536358	790	Taes_Arina_c-384354	no			
			Taes_Forno_c-805294	815	Taes_Arina_c-137542	4		*B2g20340a*	poly
			Taes_Forno_c-649793	621	Taes_Arina_c-137542	4		*B2g20340b*	poly
		Cyclin like F-box; Haem peroxidase (*B3g58810*)	Taes_Forno_c-333898	872	Taes_Arina_c-940929	9		*B3g58810b*	poly
			Taes_Forno_c-1016662	783	Taes_Arina_c-776831	1		*B3g58810a*	poly
	93B22	Cyclin like F-box; Haem peroxidase (*B3g58810*)	Taes_Forno_c-538165	296	Taes_Arina_c-776830	1		*B3g58810c*	poly
	139K17	Actin binding FH2 and DRF (*B2g38520*)	Taes_Forno_c-1027588	535	Taes_Arina_c-681963	2		*B2g38520a*	poly
			Taes_Forno_c-1084011	446	Taes_Arina_c-938299	no			
			Taes_Forno_c-1192261	300	Taes_Arina_c-938299	1			
			Taes_Forno_c-1192262	650	Taes_Arina_c-319911	2			
			Taes_Forno_c-336363	292	Taes_Arina_c-319911	1			
		Actin binding FH2 and DRF (*B2g38520*)	Taes_Forno_c-264541	325	Taes_Arina_c-66051	no			
			Taes_Forno_c-62510	4918	Taes_Arina_c-66051	11		*B2g38520b*	poly
			Taes_Forno_c-1303894	250	Taes_Arina_c-66051	no			
			Taes_Forno_c-1110189	545	Taes_Arina_c-66051	1			
			Taes_Forno_c-96618	3067	Taes_Arina_c-337529	6			
	31F17	Tpt phosphatase sugar transmembrane transporter	Taes_Forno_c-250399	1643	Taes_Arina_c-336200	no			
			Taes_Forno_c-603380	1461	Taes_Arina_c-336200	no			
		Tpt phosphatase	Taes_Forno_c-876929	308	No strong hit				
	140J13	No gene							
Ctg85	69B3	Ser/Thr phosphatase	No strong hit						
		DUF 1409; nucleic acid binding OB-fold like; Zinc finger; von Willebrand factor type A	No strong hit						
		Immunoglobulin/major histocompatibility complex	No strong hit						
		Zinc finger (*B1g75600*)	Taes_Forno_c-926238	435	Taes_Arina_c-226806	1			
			Taes_Forno_c-929319	584	Taes_Arina_c-226806	1		*B1g75600*	poly
		Glycoside hydrolase (*B1g48050*)	Taes_Forno_c-332290	3400	Taes_Arina_c-163553	1		*B1g48050*	NA
			Taes_Forno_c-441733	426	No hit				
		Pentatricopeptide repeat; Plastocyanin like (*B3g30440*)	Taes_Forno_c-584810	1400	Taes_Arina_c-748855	1		*B3g30440*	mono
	144P20	Tudor like	Taes_Forno_c-285890	273	No strong hit				
			Taes_Forno_c-1240866	1168	No strong hit				
		Tudor like (*B2g38500*)	Taes_Forno_c-1354158	679	Taes_Arina_c-410420	1		*B2g38500*	poly
			Taes_Forno_c-651137	275	Taes_Arina_c-410420	no			
		Ser/Thr phosphatase; immunoglobulin histocompatibility complex	Taes_Forno_c-106587	2180	No strong hit				
		RNA poly III (*B2g46550*)	Taes_Forno_c-550976	2076	No strong hit				

		Taes_Forno_c-638318	1123	Taes_Arina_c-56894	1	*B2g46550*	poly
		Taes_Forno_c-1279598	334	Taes_Arina_c-56894	1		
		Taes_Forno_c-80985	5053	Taes_Arina_c-112714	1		
				Taes_Arina_c-856988	2		
		Taes_Forno_c-599907	2485	Taes_Arina_c-56894	12		
	Tudor like	No strong hit					
97L13	Tudor like	Taes_Forno_c-651137	275	Taes_Arina_c-410420	no		
93G13	AMP dependent synthatase; Pentatricopeptide repeat (*B2g38490*)	Taes_Forno_c-194069	7048	Taes_Arina_c-551089	6	*B2g38490*	poly
		Taes_Forno_c-1502419	274	Taes_Arina_c-1268994	1		
		Taes_Forno_c-1371418	376	No strong hit			
134M16	AMP dependent synthatase; Pentatricopeptide repeat (*B2g38460*)	Taes_Forno_c-429451	1069	No strong hit			
		Taes_Forno_c-17366	2120	Taes_Arina_c-237016	1	*B2g38460*	poly
		Taes_Forno_c-615310	2238	No strong hit			
		Taes_Forno_c-635601	197	Taes_Arina_c-241299	1		
		Taes_Forno_c-259474	1578	Taes_Arina_c-546132	no		
110D6	Legume lectin beta domain (*B3g19870*)	Taes_Forno_c-412683	8102	Taes_Arina_c-218257	32	*B3g19870b*	NA
		Taes_Forno_c-758732	298	No strong hit			
25E20	LRR2 (*B2g16180*)	Taes_Forno_c-166738	4698	Taes_Arina_c-4840	5	*B2g16180a*	mono
		Taes_Forno_c-375463	277	Taes_Arina_c-733149	1	*B2g16180b*	mono
		Taes_Forno_c-52435	357	No strong hit			
		Taes_Forno_c-235916	1279	No strong hit			

NA: no amplification on 'Forno' and 'Arina'. Poly/mono: the markers which are either polymorphic or monomorphic between 'Arina' and 'Arina*LrFor*'.

As described in chapter 2 of this thesis, the backcross line, 'Arina*LrFor*' was initially selected for two genes, *Lr75* and *QLr.sfr-7BL*. With the aim of high resolution genetic and physical mapping of *Lr75,* the same subset of 65 NILs derived from the cross of 'Arina x Arina*LrFor*' which did not contain the 7BL QTL and which were used for genetic mapping in chapter 2, were further analyzed. The fine mapping of *Lr75* (as described in chapter 3) placed this gene towards the distal end of chromosome 1BS between markers *B2g38480* and *B2g38540a* distally and proximally, respectively. In total, 9 recombinants were identified between the two markers *B2g38480* and *B2g38540a*. As expected, all the new 15 polymorphic SNP markers mapped in the *Lr75* target interval between the two flanking markers. The 8 SNP markers *B2g38520b*, *B2g38520a, B2g20340b*, *B2g20340a*, *B2g38560, B2g38570, B2g38540b* and *B2g38540a* were co-segregating and mapped at 0.95 cM proximal to *Lr75*. Another set of proximal co-segregating markers were *B3g58810b* and *B3g58810c* which mapped 0.63 cM proximal to the gene. Both these markers were designed from the same protein Cyclin-like F-box. Surprisingly, the KASP marker *B3g58810a* though designed from the same protein as the above two, mapped 0.79 cM distal of *B3g58810b*. Towards the distal end, there were two pairs of co-segregating markers. The first pair of markers, *B2g38500* and *B2g46550*, was positioned at 0.32 cM distal to *Lr75.* Their co-segregation might be due to the fact that both these markers were designed from the same BAC clone, 144P20. The second pair of markers, *B2g38460* and *B2g38490* which co-segregated with *B2g38480* were placed at 0.48 cM distally from the gene. The markers, *B2g38480* and *B2g38460* were designed from the same BAC clone, 134M16 and the third marker, *B2g38490* from BAC clone 93G13. Both these BAC clones were assembled next to each other and co-segregation of these three markers might be due to the reason that these three markers were designed from the overlapping region between the two BAC clones. Therefore, the addition of 15 SNP markers resulted in narrowing down of the target region from 1.2 cM to 0.64 cM with *B3g58810a* and *B1g75600* as the new distal and proximal flanking markers, respectively (Fig. 4.3a, b). There is one recombinant between the distal flanking marker (*B3g58810a)* and *Lr75* whereas three recombinants were identified between the gene and proximal flanking marker (*B1g75600*). Because there is only one recombinant between the distal flanking marker and *Lr75*, we extended the target region to the cluster of three co-segregating markers (*B2g38480, B2g38460* and *B2g38490*) as the distal flanking markers. This resulted in the establishment of a *Lr75* target region of 0.96 cM.

In addition, we also wanted to map the *S/TPK_DUF2* gene because this gene was not anchored on the NIL genetic map. This is due to the reason that no SNP was identified between 'Forno' and 'Arina' from the *Serine/threonine protein kinase* gene. In order to map this gene in the NIL population, we designed a PCR based marker (*STK11*) from the sequence polymorphism between 'Forno' *S/TPK_DUF2_3* gene and *S/TPK_DUF2_1* gene of 'Chinese Spring'. Mapping of *STK11* placed it towards the distal side of *Lr75* and this maker is co-segregating with *B2g46550* and *B2g38500* markers (Fig. 4.3b).


a
Wheat chromosome 1BS physical map of cv. Chinese Spring
b
High resolution genetic map Arina x Arina*LrFor*
c
723 kb physical region of cv. Chinese Spring
d
406.3 kb physical region of cv. Forno
e
High resolution genetic map Arina x Arina*LrFor*
Scaffold 53 3.0 Mb region
Ctg_85
Ctg_638
Ctg_542
Ctg_204
Ctg_677
25E20
110D6
125P23
134M16
93G13
2H20
97L13
144P20
69B3
110J10
93B22
112H14
139K17
31F17
140J13
138P8
19B9
5L12
14P19
11M16
23A16
16A17
7I7
68G23
0.00
0.16
0.32
0.48
0.96
1.11
1.43
B2g38460
B2g38490
B2g38480
B2g38500
B2g46550
Stk11
B3g58810a
Lr75
B1g75600
B3g58810b
B3g58810c
B2g20340a
B2g20340b
B2g38520b
B2g38520a
B2g38560
B2g38570
Bg38540b
B2g38540a
~ 37 kb
~ 6.1 kb
~ 278 kb
~ 34 kb
~ 43 kb
~ 0.1 kb
~ 93 kb
~ 49 kb
~ 0.2 kb
~ 66 kb
~ 0.4 kb
Pu2
~ 35 kb
PPTR
~ 1.1 kb
AMP_dep_syn
CTREPS1
~ 273 kb
Tud_like
~ 31 kb
RP3
~ 1.7 kb
S/TPk_DUF2_1
~ 10.1 kb
Fbox_b
~ 65.4 kb
DUF_635b
~ 29.2 kb
DUF_635a
~ 42.1 kb
ZnF
~ 3.85 kb
Fbox_a
~ 3.8 kb
RN_III
~ 50.6 kb
Pu1
~ 8.3 kb
SRP54
96M10
123B20
63O11
3C21
21I24
12E6
~ 217 kb
Gap 1
~ 53.8 kb
~ 1.8 kb
~ 14.7 kb
~ 22 kb
~ 12 kb
S/TPk_DUF2_3
S/TPk_DUF2_2
Fbox
Rpb_1

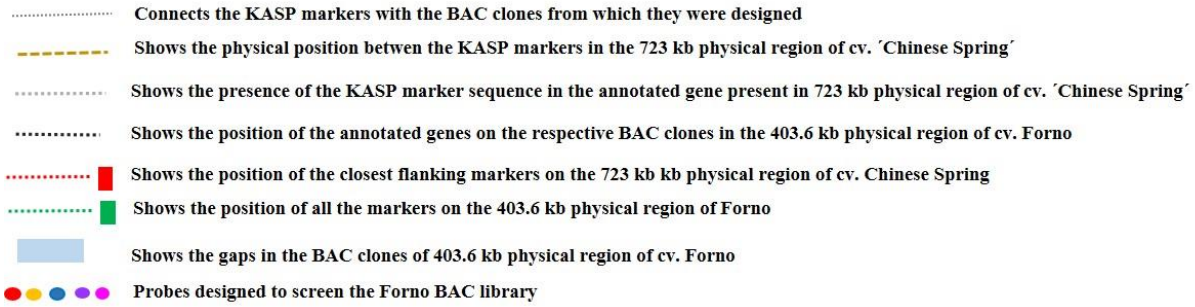


Fig. 4.3 Physical and genetic maps of wheat chromosome 1BS spanning the *Lr75* region. Panel 'a' shows the physical map of chromosome 1BS of wheat cv. 'Chinese Spring' spanning the initial 3.0 Mb region. The BAC clones marked in red represent the 723 kb physical region. Panel 'b' and 'e' are identical and show the high resolution genetic map of the 'Arina x Arina*LrFor*' mapping population. Relative distance between the markers is shown in cM on the left and right of the linkage map in panel 'b' and 'e', respectively. Panel 'c' shows the 723 kb physical region of cv. 'Chinese Spring' with the physical distances between the markers and gene annotations along with the physical distances between the genes towards left and right, respectively. Panel 'd' shows the 403.6 kb physical region of cv. 'Forno' along with the overlapping BAC clones of the 'Forno' 1X non-gridded BAC library. Physical distances are not drawn to scale. The genes marked in blue and red are the genes identified in the physical target regions of 'Chinese Spring' and 'Forno', respectively. The genes were annotated using *Brachypodium* CDS database. Red, yellow, pink and purple coloured circles show the IB-specific probes designed from the respective BAC clone for chromosome walking to cover the gaps between the BAC clone sequences in 'Forno'. *SRP54*: *Signal recognition particle 54*; *Pu1, Pu2*: *Putative1* and *Putative2*; *RN_III*: *Ribonuclease III*; *Fbox*: *Cyclin like Fbox*; *ZnF*: *Zinc finger nucleases*; *DUF_635a* and *DUF_635b*: *Domain of unknown function 635*; *STPK_DUF2_1*, *STPK_DUF2_2*, *STPK_DUF2_3*: *Serine/threonine protein kinase fused with the domain of unknown function 2*; *RP3*: *RNA polymerase III*; *Tud_like*: *Tudor like*; *AMP_dep_syn*: *Adenosine monophosphate dependent synthetase*; *PPTR*: *Pentatricopeptide repeat*; *Rpb_1*: *RNA polymerase Rpb1*.

4.4.4 Sequence characterization of the 723 kb physical target region of cv. 'Chinese Spring'

After the integration of the 15 SNP markers in the genetic map, the *Lr75* target region was narrowed down to 0.96 cM with the co-segregating cluster of three markers, *B2g38480, B2g38460* and *B2g38490* as distal and *B1g75600* as proximal flanking markers. The physical region defined by the flanking markers was comprised of 8 BAC clones (93B22, 110J10, 69B3, 144P20, 97L13, 2H20, 93G13 and 134M16, marked in red in Fig. 4.3a). After individual sequencing of these BAC clones, the 90 BAC sequence contigs present in these 8 BAC clones were manually assembled to one sequence contig of 723 kb in size. The orientation of the individual sequence contigs was determined by using the TE sequences based on the TREP database (Table 4.3). This resulted in the positioning of 34 out of 90 BAC sequence contigs. Approximately 136 kb of the sequence was only relatively positioned and oriented, meaning that these sequence contigs had overlap with TEs connecting them either to other sequence contigs of the same BAC clone or with sequence contigs of other BAC clones. The ultimate orientation and position for these remained unclear. For 21 BAC sequence contigs, amounting to a sequence size of ~142 kb, no overlaps were identified and therefore they were integrated in the forward orientation at a random position. For BAC clone 2H20 only 3 out of 24 sequence contigs were integrated into the map, since those 3 sequence contigs were already sufficient to cover the whole BAC clone insert. Finally, a total of 69 BAC sequence contigs from the 8 BAC clones were assembled to a one continuous BAC sequence contig of 723 kb in size (Table 4.3). The order of the 8 overlapping BAC clones in the target interval from the proximal to the distal end is depicted in the order as described in Table 4.3 from top to bottom.

Table 4. 3 Overview of assembled BAC clone sequence contigs merged to form one continuous physical sequence of 723 kb

BAC clone	**Total ctgs**	**positioned and orientated sequences**		**relatively positioned and orientated sequences**		**not orientated and randomly positioned sequences**	
		No. of contigs	**Size in bp**	**No. of contigs**	**Size in bp**	**No. of contigs**	**Size in bp**
93B22	15	8	75,503	7	58,451	-	-
110J10	10	4	74,066	4	58,296	2	1,866
69B3	9	5	62,738	-	-	4	25,761
144P20	11	4	92,917	2	9,489	5	85,884
97L13	4	3	145,445	1	9,881	-	-
2H20	24	3	116,474	-	-	-	-
93G13	2	1	133,612	-	-	1	1,446
134M16	15	6	128,107	-	-	9	27,364
Total	90	34	828,862	14	136,117	21	142,321

4.4.5 Comparison of the genetic and physical maps

The position and order of the 15 mapped markers that were designed from the SNPs between the flow-sorted Illumina sequences of 'Forno' and 'Arina' on the 'Arina x Arina*LrFor*' genetic and 'Chinese Spring' physical maps were compared. The comparison revealed a largely collinear order of the molecular markers on both maps. Marker *B3g58810b* mapped at two different positions on the physical map, indicating a duplicated gene sequence. The recombination rate varied greatly within the target interval. The genetic distance of 0.32 cM between markers *B2g20340a* and *B3g58810b* corresponded to a physical region of 66 kb, with a physical to genetic distance ratio of 213 kb/cM. Similarly, the genetic distance of 0.64 cM between the markers, *B3g58810a* and *B1g75600* resulted in a physical to genetic distance ratio of 145.31 kb/cM (Fig. 4.3c). In contrast, the genetic distance of 0.16 cM between the distal markers, *B2g38500* and *B2g38480* corresponds to a physical distance of 278 kb with a physical to genetic distance ratio of 1.7 Mb/cM. We also observed that the individual markers of co-segregating clusters are physically separated from each other. For example, the co-segregating markers *B2g46550* and *B2g38500* were separated by a physical distance of 34 kb.

4.4.6 BAC clones spanning the *Lr75* region in leaf rust resistant cv. 'Forno'

The wheat cv. 'Chinese Spring' lacks leaf rust resistance gene *Lr75*. We therefore wanted to establish a physical region spanning the *Lr75* locus from the donor cv. 'Forno'. To achieve this, a non-gridded BAC library with 1X coverage of the 'Forno' genome was constructed in collaboration with the French Research Genomic Centre, CNRGV, France. This BAC library was screened with chromosome 1B-specific probes which were designed from those flow-sorted Illumina sequence contigs of cv. 'Forno' that aligned with gene containing sequences present in the 723 kb physical sequence contig of cv. 'Chinese Spring'. Three probes, Fbox2F/4R (designed from the *Fbox* gene coding sequence) and STK1F/2R (designed from *Serine/threonine protein kinase* gene coding sequence) and B38460.3 (from *Putative 2* gene coding sequence) were used initially to screen the 'Forno' BAC library (Fig. 4.3c, orange circles). The PCR pool screening of the 'Forno' BAC library with the three probes, Fbox2F/4R, STK1F/2R and B38460.3 resulted in the identification of 3 BAC clones, 12E6 (80kb), 3C21 (95kb) and 96M10 (95 kb), respectively. 'Chromosome walking' was performed to cover the gap between the three BAC clones of cv. 'Forno'. This was achieved by using two approaches to design more 1B-specific probes. The first approach was to design the probes from the identified gene sequences on the three BAC clones. The second approach for probe designing was the use of BES of the 'Forno' BAC clones 12E6 and 3C21. The BES were aligned against the TREP database and BES with more than 80% homology with the TE sequences were considered. The probes were designed from the regions which are present at the identified repeat sequence junctions. Using these approaches, two probes RP3M14 (Fig. 4.3d, dark blue circle) and TREP15F/18R (Fig. 4.3d, purple circle) were designed. RP3M14 was designed from the *RP3* nucleotide sequence on 12E6 BAC clone and TREP15F/18R from the repeat sequence junction between the genes *S/TPK_DUF2_2* and

S/TPK_DUF2_3 on the BAC clone 3C21. The screening of the 'Forno' BAC library with these two new probes resulted in the identification of the same BAC clone, 21I24 (94 kb), thereby covering the gap between 12E6 and 3C21. The three BAC clones span one continuous physical sequence contig of 188 kb which included the two closest flanking markers of the *Lr75* gene (*B3g58810a* and *B1g75600*). We identified one inversion while comparing the genetic map and the 188 kb physical region of 'Forno'. The inversion is in between markers *B1g75600* and *B3g58810c*. We speculate that this inversion is because of the wrong assembly of BAC clone12E6 which needs to be re-sequenced with PACBio sequencing platform in near future to resolve this problem.

To the best of our data analysis, we identified four recombinants in the *Lr75* target region between the closest flanking markers. Three recombinants were identified between *Lr75* and the proximal flanking marker (*B1g75600*) and only one recombinant was identified between *Lr75* and distal flanking marker (*B3g58810a*). However, these recombinants need to be confirmed in the 2016 field season. Therefore, we extended the physical region in 'Forno' to the cluster of three co-segregating markers (*B2g38480, B2g38460* and *B2g38490*) towards the distal end. In order to further cover the gap between the 188 kb physical sequence contig and the BAC clone 96M10, more probes were designed from the TE sequence junctions of both 723 kb physical sequence contig of cv. 'Chinese Spring' and 188 kb physical sequence contig of cv. 'Forno'. From the large gene-free region of 273 kb between the genes *Tud-like* and *AMP-dep-syn* present in the 723 kb physical sequence contig of cv. 'Chinese Spring' another probe, CTREP51 was designed (Fig. 4.3c, red circle). This probe was used to screen the 'Forno' BAC library and one additional BAC clone, 63O11 (117 kb) was identified. This new BAC clone lied in between the two identified regions but did not cover the gap from either side. BES of the new BAC clone 63O11 were used and a new 1B-specific probe, TREP61F/62R was designed from the repeat sequence junctions towards the distal end (Fig. 4.3d, pink circle on 63O11 BAC clone). With this probe, a new BAC clone 123B20 (106 kb) was identified which overlaps with 63O11 and 96M10 BAC clones. So far, the BAC library screening of cv. 'Forno' resulted in approximately 403.6 kb physical sequence region with one gap (Fig. 4.3d, light blue horizontal bar) between the BAC clones 63O11 and 3C21 which needs to be covered by screening the 'Forno' BAC library with additional 1B-specific probes.

4.4.7 Gene annotation of the cv. 'Chinese Spring' and cv. 'Forno' target region for candidate gene identification

The gene annotation of the 723 kb physical sequence contig of cv. 'Chinese Spring' using the *Brachypodium* CDS database resulted in the identification of 14 genes between the markers *B2g38460* and *B2g20340b* (Fig. 4.3b, c). Two genes i.e. *Cyclin-like F-box* and *DUF635* were duplicated in the 723 kb physical sequence contig. The other genes present in this region were *Serine/threonine protein kinase* fused with a *domain of unknown function, RNA polymerase III, Pentatricopeptide repeat, AMP*

dependent synthetase, *Zinc finger*, *Signal recognition particle 54* and two more putative genes which were named as *Putative 1* and *Putative 2* (Fig. 4.3c). Table 4.4 describes all the identified genes in the target interval. The *Serine/threonine protein kinase* gene has a chimeric structure where the two domains, a serine/threonine kinase domain (S/TPK) and a domain of unknown function (DUF2) were fused together. The S/TPK domain of cv. 'Chinese Spring' shared 96% nucleotide sequence similarity with the homolog in *Brachypodium*, and the DUF2 domains shared 85% sequence similarity between 'Chinese Spring' and *Brachypodium*. By comparing the nucleotide sequence of the DUF2 domain of wheat with barley, it was found that the closest homolog in barley is Hvul_AK372571, which shared 84% sequence similarity. The fusion of S/TPK and a second domain of unknown function (DUF2) reported in this study is also unique. Comparison of the unique gene sequence with the closest relative genome of *Brachypodium* as well as with the genomes of barley, rice and *Arabidopsis*, showed that this unique fusion might have specifically evolved in the *Brachypodium*/Triticeae lineage. The *Brachypodium* genome also contains this unique combination, which is nearly 85% identical to that of wheat. In contrast, *Arabidopsis* and other grasses possess 2-3 *DUF1-DUF2* genes i.e. the DUF2 domain is present in combination with different DUF domains. Interestingly, genes with the combination of the DUF1-DUF2 domains are also present in *Brachypodium* and Triticeae along with the presence of multiple copies of genes encoding S/TPK-DUF2 type of proteins.

The nucleotide sequences of the two *Fbox* gene copies are 100% identical in 'Chinese Spring'. The two genes are in opposite orientation and there is 97% similarity between the 3' and 5' UTRs of *Fbox_a* and *Fbox_b* genes. Furthermore, comparisons of the gene sequences of 'Chinese Spring' with the flow-sorted sequences of susceptible cv. 'Arina' revealed 100% similarity, revealing that 'Arina' haplotype is very close to 'Chinese Spring'.

Annotation of the BAC clones in the approximately 403.6 kb 'Forno' physical sequence contig resulted in the identification of one additional gene, *RNA polymerase Rpb1*, which is not present in the 'Chinese Spring' physical sequence contig. In total, the annotation identified 8 genes in the *Lr75* physical sequence contig of cv. 'Forno'. Of them, *Fbox* is the only gene which is present between the closest flanking markers (*B3g58810a* and *B1g75600*) of *Lr75* gene (Fig. 4.3d). Therefore, this gene is the most promising putative candidate gene. Because only one recombinant was identified between *Lr75* and *B3g58810a*, we extended the target region to a set of three co-segregating markers *B2g38480, B2g38460* and *B2g38490* towards the distal end. This led to the identification of two other genes, *S/TPK_DUF2_2* and *S/TPK_DUF2_2* as putative candidate genes. The genes are listed in Table 4.4.

Table 4.4 Gene annotation of the complete 'Chinese Spring' *Lr75* physical sequence contig and its comparison with the corresponding partial *Lr75* physical sequence contig in 'Forno'

No.	Name	Annotation	*Brachypodium* gene	CS Gene start	CS Gene end	Orientation in CS	Forno Gene start	Forno Gene end	Orientation in Forno
1	*Rpb1*	*RNA polymerase Rpb1*	*Bradi5g_23820*	-	-		4317	6310	reverse
2	*SRP54*	*Signal recognition particle 54*	*Bradi2g_20340*	43752	46126	forward	-	-	
3	*Pu1*	*Putative gene 1*	*Bradi2g_18390*	54347	54922	reverse	-	-	
4	*RNIII*	*Ribonuclease III*	*Bradi2g_52920*	105570	105774	reverse	-	-	
5	*Fbox*	*Cyclin like Fbox*	*Bradi3g_58810*	109593	111196	forward			
				254408	256010	reverse	40831	41691	forward
6	*ZnF*	*Zinc Finger*	*Bradi5g_07340*	115050	115517	reverse	-	-	
7	*DUF635a*	*Domain of unknown function*	*Bradi1g_54870*	157631	158683	reverse	-	-	
8	*DUF635b*	*Domain of unknown function*	*Bradi1g_54870*	187925	189025	forward	-	-	
9	*S/TPK_DUF2*	*Serine /Threonine protein kinase_Domain of unknown function*	*Bradi2g_38510*	266076	278964	forward			
							82231	99788	reverse
							153637	160128	reverse
10	*RP3*	*RNA polymerase III*	*Bradi2g_46550*	280627	300168	forward	56318	80410	reverse
11	*Tud_like*	*Tudor like*	*Bradi2g_38500*	331148	335963	reverse	*	*	
12	*AMP_dep_syn*	*AMP dependent synthetase*	*Bradi2g_38490*	608897	614287	forward	333460	338843	forward
13	*PPTR*	*Pentatricopeptide repeat*	*Bradi2g_38480*	615409	617322	Reverse	339520	341433	reverse
14	*Pu2*	*Putative gene 2*	*Bradi2g_38470*	652407	653930	reverse	346410	347915	reverse

- Indicates a missing gene in the respective haplotype. * indicates that the gene has not been found in 'Forno' but might still be present in the gap

4.4.8 Haplotype comparison between 'Chinese Spring' and 'Forno' physical target intervals

Haplotype comparisons between 'Chinese Spring' and 'Forno' in the physical target region revealed considerable differences with inversions, deletions and duplications. Small regions of high similarity with some re-arrangements between the two haplotypes were evident towards the proximal end. In contrast, the distal part of nearly 200 kb has a high sequence similarity between 'Chinese Spring' and 'Forno'. The physical sequence contig of 'Forno' has one gap, Gap 1 which is at 27.46 kb distal to the *S/TPK_DUF2_3* gene (Fig. 4.4, red bar). This gap needs to be covered in the future with additional BAC clone sequences by screening the 'Forno' BAC library. A region of nearly 130 kb in 'Chinese Spring' has no sequence conservation with 'Forno'. This region also contains the *Tud_like* gene in 'Chinese Spring' which is not present in 'Forno'. The divergence of this 130 kb region in both haplotypes might also be partially due to the gap in the 'Forno' sequence. It is possible that the T*ud_like* gene is actually not deleted in 'Forno' but is present in the gap.

Both haplotypes show deletion of genes. The *Rpb1* gene present in 'Forno' is deleted in 'Chinese Spring' but the region around this gene is conserved in both haplotypes. Interestingly, a set

of duplicated genes, *DUF635a* and *DUF635b* is deleted in 'Forno' but a high sequence conservation of approximately 18 kb between the two genes was evident between two haplotypes. Further, the *Fbox* gene is duplicated in 'Chinese Spring' whereas, only a single copy is present in 'Forno'. A duplication of the *S/TPK_DUF2* gene is observed in 'Forno' as compared to the single copy in 'Chinese Spring'. A total of three inversion events was detected between the two haplotypes. A first inversion of 5 kb occurs at the region of *Fbox_b* gene. A second and third inversion occur in the region of *RP3* gene sequence.

Comparison of the *Fbox* gene of 'Forno' revealed highest homology with *Fbox_b* of 'Chinese Spring'. The *Fbox* gene in 'Chinese Spring' has a CDS sequence of 1602 bp whereas in 'Forno' this gene has a size of 861 bp which encodes a putative polypeptide of 287 amino acids. The *Fbox* gene in 'Forno' is an inactive pseudogene as the initial 741 bp of coding sequence is deleted. The nucleotide sequences of the two S/TPK domains in 'Forno' showed 94% sequence similarity whereas, the DUF2 domains showed 86% similarity, therefore indicating a high sequence conservation of S/TPK domains. Overall, there is 85-99% nucleotide sequence identity between the predicted genes of the two haplotypes. Despite the re-arrangements between the two haplotypes, six genes, *Fbox, RP3, S/TPK_DUF2, AMP_dep_syn, PPTR* and *Pu2* are present in both (Fig. 4.4).

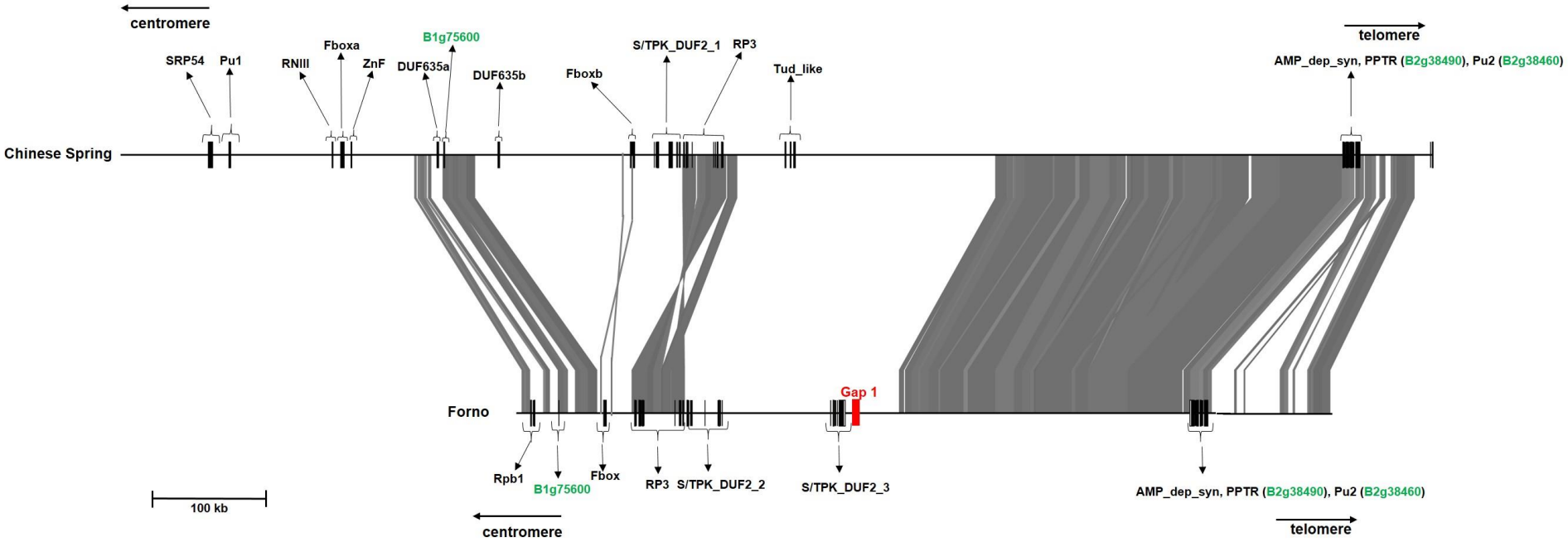


Fig. 4.4 Haplotype comparison between the 723 kb physical target sequence of cv. 'Chinese Spring' and approximately 403.6 kb physical target sequence of cv. 'Forno'. Vertical black bars indicate the position of exons. Curl brackets group the exons of each gene and the gene names are shown with arrowheads. Position of exons is drawn to scale. Red bar indicates the gap in the 'Forno' physical sequence. *SRP54*: *Signal recognition particle 54*; *Pu1*, *Pu2*: *Putative1* and *Putative2*; *RN_III*: *Ribo nuclease III*; *Fbox*: *Cyclin like Fbox*; *ZnF*: *Zinc finger nucleases*; *DUF_635a* and *DUF_635b*: *Domain of unknown function 635*; *S/TPK_DUF2_1*, *S/TPK_DUF2_2*, *S/TPK_DUF2_3*: *Serine/threonine protein kinase fused with the domain of unknown function 2*; *RP3*: *RNA polymerase III*; *Tud_like*: *Tudor like*; *AMP_dep_syn*: *Adenosine monophosphate dependent synthetase*; *PPTR*: *Pentatricopeptide repeat*; *Rpb_1*: *RNA polymerase Rpb1*. Markers in green are the proximal and distal flanking markers and define the *Lr75* target region of 0.96 cM.

4.5 Discussion

4.5.1 LTC assembly of BAC contigs improves quality of physical mapping

Manual sequence alignment of the first set of 24 BAC clones obtained through FPC assembly resulted in gaps in the physical target region of 'Chinese Spring'. By using the new BAC clone information obtained from LTC assembly (Frenkel et al. 2010), the gaps in the physical region could be closed. Thus, our results show the advantage of LTC assembly over FPC assembly. Raats et al. (2013) also discussed the advantage of LTC over FPC in their study. They obtained more reliable and longer BAC contigs by LTC than by FPC assembly with a largest contig size of 7.0 Mb with LTC as compared to 4.7 Mb with FPC. However, the coverage of chromosome 1BS was slightly higher (88%) for FPC than for LTC (87%). This was due to the presence of nearly 1500 problematic clones which are either poorly fingerprinted or are chimeric. They called such clones questionable clones (Q-clones) in the FPC assembly, but they were removed from the LTC assembly. Also, by using LTC, Raats et al. (2013) merged 254 BAC contigs into 57 long scaffolds. Breen et al. (2013) while working on the physical mapping of wheat chromosome 1AS also compared the LTC assembly with the FPC assembly. In their study, they generated 394 LTC contigs as compared to 805 contigs generated by FPC assembly. In order to compare the LTC and FPC contigs, Breen et al. (2013) also designed software with which they produced 208 LTC contigs as a result of the fusion of 472 FPC contigs which were not significantly overlapping according to FPC assembly. The software generated by Breen et al. (2013) also showed the higher reliability of LTC assembly due to the production of larger and integrated assemblies as compared to the FPC one. Therefore, all the studies that used LTC assembly produced longer contigs with better ordering. This led to better selection of MTP BAC clones with less number of Q-clones and hence development of better and more reliable physical maps as compared to FPC assembly.

4.5.2 The leaf rust resistance gene *Lr75* might be different from the class of NBS-LRR resistance genes

The classical race-specific resistance (*R*) genes mainly encode for protein immune receptors in plants that interact with fungal effectors or avirulence (*Avr*) proteins, resulting in hypersensitive response (HR). Most of the disease resistance genes cloned so far in wheat, such as *Lr1* (Cloutier et al. 2007), *Lr10* (Feuillet et al. 2003), *Lr21* (Huang et al. 2003); stem rust resistance genes, *Sr33* (Periyannan et al. 2013), *Sr35* (Saintenac et al. 2013a); stripe rust resistance gene, *Yr10* (Liu et al. 2014); and powdery mildew resistance genes, *Pm3b* (Yahiaoui et al. 2004), *Pm8* (Hurni et al. 2013) belong to the NBS-LRR class. All these genes are effective at seedling stage and most of these genes, except for *Sr33* and *Sr35*, are race-specific, i.e. they are effective against a particular set of pathogen races. Both stem rust resistance genes mentioned above confer resistance to various stem rust races including Ug99 group. In contrast to these seedling resistance genes, the phenotypic evaluation of

Lr75 in the field (described in chapter 2 of this thesis) showed that *Lr75* is a partial adult plant leaf rust resistance gene. Interestingly, the 723 kb physical target interval spanning the *Lr75* locus from the hexaploid wheat cv. 'Chinese Spring' as well as the partial physical target region from the resistant winter wheat cv. 'Forno' do not contain any NBS-LRR genes. Therefore, we speculate that the *Lr75* gene belongs to a class of resistance genes other than NBS-LRR.

4.5.3 Sequence divergence between haplotypes

Several insertion-deletion as well as gene duplication events indicate divergence in the *Lr75* region between the 'Forno' and 'Chinese Spring' haplotypes. We identified a duplication of a *Serine/Threonine protein kinase (S/TPK)* gene fused with a domain of unknown function (*DUF2*) in the physical sequence contig of resistant cv. 'Forno'. Cv. 'Chinese Spring' has only one copy of the *S/TPK_DUF2* gene. We also observed the deletion of *DUF635a* and *DUF635b* genes in 'Forno'. There is one gap of unknown size in the 'Forno' sequence which needs to be covered by screening the 'Forno' BAC library for additional overlapping BAC clones. Due to poor sequence conservation between 'Chinese Spring' and 'Forno' in the region of Gap1, we cannot predict the size of Gap 1 in 'Forno'. Such large insertion-deletions had been reported previously in wheat and other cereals for other loci. For instance, Mago et al. (2014) observed a deletion of 10 *Germin-like Protein* (*GLP*) genes in a physical region of cv. 'Chinese Spring' which is susceptible to stem rust as compared to the resistant cv. 'Hope' that carries the *Sr2* gene. Also for stripe rust APR gene *Yr36*, a 183 kb insertion was present in resistant line RSL65 in comparison to the susceptible durum cv. 'Langdon' in wheat (Fu et al. 2009). There, the deleted region consisted of three genes, *in between RING finger domain* (*IBR1*) and a pair of duplicated *WHEAT KINASE START 1* and *WHEAT KINASE START 2* genes (*WKS1* and *WKS2*). Similarly, in *Triticum dicoccoides* as compared to cv. 'Langdon', a deletion of a more than 300 kb region was reported which contains a toxin (ToxA) sensitivity gene, *Tsn1* for tan spot disease in wheat (Faris et al. 2010; Faris et al. 2013). Despite the availability of a draft genome sequence of cv. 'Chinese Spring' of wheat, map-based isolation of genes of agronomical importance will be hampered by such differences between different wheat haplotypes. Therefore, it is essential to get the complete physical target region from the donor cultivars. Generating a non-gridded BAC library of the resistant cultivar is a good alternative to identify candidate genes for map-based cloning projects. The non-gridded BAC library of cv. 'Forno' generated during this project revealed divergence between 'Forno' and 'Chinese Spring' and therefore, confirmed the importance of a BAC library from the resistant donor parent. Flow-sorting of individual chromosome arms (Molnár et al. 2014) further reduces the complexity of genetic and genomic analysis and can lead to the faster establishment of physical regions from the resistant cultivars spanning the gene of interest. Using this technique, only the chromosome containing the gene of interest from the resistant cultivar can be

sorted and a BAC library can be generated for the flow-sorted chromosome to get the physical target region.

4.5.4 Candidate genes identified in the 'Forno' physical target region

So far 8 genes have been identified in the physical target region of cv. 'Forno'. Out of these 8 genes, *Fbox* gene is the most promising candidate because it lies between the closest flanking markers in the genetic map. Other than the *Fbox* gene, two other genes i.e. *S/TPK_DUF2_2*, *S/TPK_DUF2_3* are also promising candidates because of the role of similar proteins in the plant defense response against plant pathogens (Afzal et al. 2008). The *S/TPK_DUF2* gene is a fusion product of the S/TPK domain with an unknown domain, DUF2 and represents a chimeric gene.

Genes encoding cyclin-like F-box proteins represent one of the largest multigene families. F-box proteins contain a 40-60 amino acid domain at the N-terminus which is called the F-box. They are the main component of the Skp1p-cullin-F-box (SCF) complex of E3 ubiquitin ligase, which is formed by four different proteins: Skp1, Cullin, Rbx1 and F-box. This complex is part of the ubiquitin-proteasome pathway which regulates protein degradation by the 26S proteasome mediated by the action of E1, E2 and E3 enzymatic complexes. The SCF complex plays a crucial role in cellular processes such as biotic and abiotic stress response, plant growth and development, hormonal response and floral development (Smalle and Vierstra 2004; Dreher and Callis 2007; Vierstra 2009). It is also well documented in the literature that several plant-pathogenic microbes interact with the components of the SCF complex via F-Box motifs and lead to increased infection by manipulating the ubiquitin-proteasome pathway (Banks et al. 2003). The pathogens either directly encode the component of E3 enzymes or host E3 enzyme components are redirected by the pathogen (Lechner et al. 2006). One such example is the interaction of the *Arabidopsis* infecting polerovirus P0 with the Skp1 component by the F-Box motif (Pazhouhandeh et al. 2006). Posttranscriptional gene silencing mediated by small RNA is a strategy employed by plants in order to defend themselves against viral attack. In response to this, plant viruses encode for proteins that can suppress this gene silencing. Pazhouhandeh et al (2006) proposed a model where P0 virus encodes an F-Box protein which targets the posttranscriptional gene silencing of the host plant, thereby rendering the plant susceptible. A mutation in the F-Box motif blocked the P0-Skp1 interaction, thereby reducing the suppressor activity and making the plant more resistant to the virus. In our study we also found an F-Box gene whose N-terminal is truncated in the resistant parent 'Forno'. In contrast, 'Chinese Spring' which lacks *Lr75* has a complete open reading frame (ORF) of the F-Box protein in the target region. Thus, it is possible that proteins of the leaf rust fungus bind to the F-box motif and interact with the SCF complex, making the plant susceptible. In this model, the truncation of the N-terminus F-Box protein would interrupt the fungal interaction with the SCF complex and make the plant more resistant. It is quite possible that due to the loss-of-function mutation in the *F-box* gene, it might be a recessive resistance gene. However, so far no information is available on the dominance/recessive nature of

Lr75. Screening of the F_1 plants against leaf rust at the adult plant stage in the field will give a clear answer to the dominance/recessive nature of *Lr75*.

4.5.5 Evolution of chimeric genes

The annotation of a physical region of cultivars 'Chinese Spring' and 'Forno' identified several copies of *Serine/threonine kinase* genes fused with a *domain of unknown function (DUF2)*, thereby, forming a chimeric gene structure which we designated as, *S/TPK-DUF2_1*, *S/TPK_DUF2_2* and *S/TPK_DUF2_3*. The S/TPK domains show more than 90% nucleotide sequence similarity whereas the DUF2 domains are nearly 85% identical. Previous studies also reported such chimeric genes involved in resistance. The previously cloned broad-spectrum, adult plant stripe rust resistance gene, *Yr36* encodes the WKS1 protein which is a fusion of a kinase domain and a START domain (Fu et al. 2009). This fusion is unique and no such type of domain fusion has been reported in other organisms. The stem rust resistance gene *Rpg5* in barley also has a unique combination of two domains (Brueggeman et al. 2008). The gene *Rpg5* encodes an N-terminal nucleotide binding site-leucine rich repeat (NBS-LRR) domain and a C-terminal serine/threonine protein kinase (S/TPK) domain. It was found that both domains are necessary for resistance to stem rust. In this study it was also postulated that the LRR domain acts as a pathogen receptor, while both NBS and S/TPK domains are responsible for signal transduction. In contrast, a fungal toxin (ToxA) sensitivity gene, *Tsn1* in wheat, which provides susceptibility to tan spot and Staganospora nodorum blotch toxins, encodes a fusion protein of an S/TPK domain at the N-terminus and a NBS-LRR domain at the C-terminus (Faris et al. 2010). The authors reported that all three domains are required for sensitivity to ToxA and disease susceptibility. They also showed through phylogenetic analysis that the evolution of this unique combination was due to a gene fusion event in the B-genome progenitor of polyploid wheat. Both *Tsn1* and *Rpg5* genes possess the same three domains i.e. NBS, LRR and S/TPK in a single transcript but in different arrangements. Faris et al. (2010) also reported that both genes arose through two independent gene fusion events and therefore, do not share any common ancestry.

The fusion of the S/TPK protein with other domains has not only been observed in proteins active against fungal pathogens, but also against bacterial pathogens. For example, a disease resistance gene in rice, *Xa21* that provides resistance to the bacterial blight pathogen (*Xanthomonas oryzae* pv. o*ryzae*) has both an S/TPK domain and an LRR domain. Both domains play a role in cell surface recognition of the pathogen and finally activate the intracellular plant defense response (Song et al. 1995). Similarly, in *Arabidopsis*, the gene *FLS2* confers sensitivity to bacterial flagellin elicitor flg22, thereby inducing the defense response in plants (Gómez-Gómez and Boller 2000). The gene *FLS2* also possesses the unique structure with a combination of LRR and S/TPK domains.

Other genes have been cloned in plants where both the NBS-LRR and S/TPK domains are required for disease resistance, but the two domains are not present in a single protein. For instance, *Pto* gene in Tomato and *PBS1* gene of *Arabidopsis* encode an S/TPK domain; and *Prf* and *RPS5* in tomato and *Arabidopsis*, respectively encode NBS-LRR domains (Salmeron et al. 1996; Swiderski and Innes 2001).

4.6 Outlook

For the identification of all possible candidate genes, a complete physical region spanning *Lr75* from 'Forno' has to be established. Furthermore, for the validation of candidate genes, a set of putative M4 EMS mutants were selected. These mutants have the 'Forno' genotype for both *Lr75* and *QLr.sfr-7BL* QTL regions. The complete coding sequences of the possible candidate genes will be amplified in the mutants by using the probes designed from the genes. For validation of the candidate genes, they will then be transformed into the wheat cv. 'Forno' using stable transformation assay.

Chapter 5

General Discussion and Outlook

5.1 Race-specificity of 'Arina*LrFor*' at the seedling stage

5.1.1 Suppression of resistance by the presence of suppressor genes

5.1.2 Putative interaction of *Lr75* or *QLr.sfr-7BL* with a seedling resistance gene in 'Arina' background

5.2 Origin of *Lr75* in the wheat germplasm

5.3 High throughput genotyping in wheat

5.3.1 Exome capture as an efficient tool for targeted-capture resequencing of parts of wheat genome

5.3.2 SNP genotyping in wheat

5.4 Cloning of *Lr75* and approaches for validation of candidate genes

5.1 Race-specificity of 'Arina*LrFor*' at the seedling stage

Lr75 is a partial, adult plant leaf rust resistance gene. It confers a slow-rusting resistance response on the flag leaves at the adult plant stage in the field. Partial adult plant resistance genes mostly do not show a resistance response at the seedling stage. But to our surprise, we observed a resistance response of the backcross line 'Arina*LrFor*' that carries both *Lr75* and *QLr.sfr-7BL* genes at the seedling stage against four of the seven tested leaf rust isolates. However, the cv. 'Forno' which is a donor for the two QTLs, *Lr75* and *QLr.sfr-7BL* showed susceptible response to all the leaf rust isolates tested so far. This led to two hypotheses: 1) there might be a suppressor gene present in 'Forno' which suppresses *Lr75* and *QLr.sfr-7BL* gene activity at the seedling stage or 2) *Lr75* or *QLr.sfr-7BL* might interact and enhance the effect of a seedling resistance gene present in the genetic background of 'Arina' but absent in 'Forno'.

5.1.1 Suppression of resistance by the presence of suppressor genes

The suppression of resistance genes is a common phenomenon in plant breeding. In wheat breeding, resistance genes introgressed into hexaploid wheat cultivars from wheat relatives with lower ploidy level or wild wheat relatives, often show reduced effectiveness to disease or are suppressed by the presence of suppressors (Mcintosh et al. 2011). Well described examples are the stem rust resistance gene *Sr21* which is originally present in diploid wheat *Triticum monococcum* and leaf rust resistance gene *Lr23* present in tetraploid wheat, *Triticum turgidum* (The and Baker 1975; Nelson et al. 1997). When introduced into hexaploid wheat, their effects got suppressed by the presence of suppressor genes. Kerber and Green (1980) in their study also made a similar observation where the tetraploid wheat cultivar 'Tetra Canthatch' (2n = 28 = AABB) has shown higher resistance to stem rust (*Puccinia graminis* f. sp. *tritici*) races as compared to the hexaploid wheat cv. 'Canthatch'. Suppression of resistance to all the three rusts has been well studied in wheat germplasm (McIntosh and Dyck 1975; Bai and Knott 1992; Williams et al. 1992; Kema et al. 1995; Knott 2000; Chen et al. 2013; Talajoor et al. 2015; Wu et al. 2015). Suppressors have also been described for resistance to other diseases. For example, the *Pm3* gene is the suppressor of powdery mildew resistance gene *Pm8* derived from rye (Mcintosh et al. 2011; Hurni et al. 2014). Hurni et al. (2014) in their study explained for the first time the molecular basis of suppression of a resistance gene in wheat. By using a transient expression assay in *Nicotiana benthamiana*, they showed that both the suppressed (*Pm8*) and the suppressor (*Pm3*) genes are expressed and interact. They concluded that gene silencing or mutation or gene loss were not the cause of suppression of *Pm8* resistance gene. Instead, the suppression of *Pm8* is most likely due to a post-translational mechanism such as non-functional heteromeric protein complex formation. It was postulated in the literature that the presence of the orthologous genes on homeologous chromosomes might be the cause of suppression of resistance genes in hexaploid wheat (Nelson et al. 1997; Hurni et al. 2013). During this study, we also observed a resistance response of 'Arina*LrFor*' against some leaf rust isolates at the seedling stage as compared to the donor cv. 'Forno'

which shows a susceptible response against all tested isolates. We propose a hypothesis that there is a suppressor of *Lr75* or *QLr.sfr-7BL* present in 'Forno' that suppresses the effect of these genes at the seedling stage. To understand the mechanism of suppressor gene action, it is important to map the suppressor gene. To do that, we will use a set of $F_{4:5}$ derived RIL lines (containing *Lr75* and *QLr.sfr-7BL* individually without *Lr34*) which were generated from the cross of 'Arina x Forno'. If the suppressor is present in 'Forno' then we expect a segregation in the RIL lines on testing with the isolate which is virulent on 'Forno'. Comparison of the phenotypic data with the genotypic data of the RIL population constructed by Schnurbusch et al. (2004) should give a rough genetic position of the suppressor gene. For efficient introgression of resistance genes in wheat breeding, it is important to identify and map the additional suppressor genes in wheat.

5.1.2 Putative interaction of *Lr75* or *QLr.sfr-7BL* with a seedling resistance gene in the 'Arina' background

Our second hypothesis is that the resistance response of 'Arina*LrFor*' at the seedling stage against some isolates might be due to the interaction of *Lr75* or *Qlr.sfr-7BL* with another race-specific seedling resistance gene 'X' present in the 'Arina' but not in the 'Forno' background. Either *Lr75* or *QLr.sfr-7BL* might be capable of "boosting" a resistance gene in 'Arina' which is either not expressed in seedlings when present alone or for which resistance has been broken and is now reinstated by the combination with *Lr75* or *QLr.sfr-7BL*. A possible interaction partner could be the seedling resistance gene *Lr13* present in 'Arina' (Pathan and Park 2006). However, the two isolates (96002, 96209) against which 'Arina*LrFor*' showed a resistance response are virulent on *Lr13* according to the virulence/avirulence formulae of the isolates (Appendix B). Therefore, *Lr13* is unlikely to be the interaction partner in 'Arina'. So far, none of the identified leaf rust resistance genes correlates with our observations of the seedling assay. Therefore, we can speculate that race-specificity of 'Arina*LrFor*' at the seedling stage might be due the interaction of some unidentified gene in 'Arina'. To map the new gene 'X', we will use a subset of lines of the $F_{5:7}$ 'Arina x Forno' RIL population (containing *Lr75* and *QLR.sfr-7BL* individually but not *Lr34*) and infect them with the isolate for which 'Arina*LrFor*' has shown resistance at the seedling stage. The infection should result in segregation and the possibility to get a rough genetic mapping of 'X'. To map 'gene X' at higher resolution, we will make crosses between the resistant and susceptible RIL lines. Subsequent selfing will result in a mapping population which can be used to map the 'X gene' using high throughput genotyping technologies.

5.2 Origin of *Lr75* in the wheat germplasm

So far, we do not have much information about the origin of *Lr75* except that it is present in 'Forno'. The gene was not described in CIMMYT wheat germplasm or other germplasm collections. The origin of *Lr75* was first hypothesized to be in the Russian wheat germplasm, since the *Lr34* allele found in 'Forno' is suggested to be inherited from the Russian wheat varieties 'Bezostaya' and 'Kavkaz' (Krattinger et al. 2009). This possibility can be excluded as those lines are in possession of the 1RS.1BL translocation (GRIS 2014), whereas 'Forno' evidently does not belong to this class of wheat-rye translocation lines. To determine the origin of *Lr75* we genotyped a set of 43 wheat lines by using a subset of 66 SNP markers derived from the target region of *Lr75* (presented in section 3.4.5 of chapter three of this thesis). The analysis revealed higher frequencies of 'Forno' alleles in two Swiss wheat cultivars, 'Forel' and 'Fiorina'. It could be possible that these two cultivars contain *Lr75*. To confirm the presence of *Lr75* in these two cultivars, the closest *Lr75* flanking markers will be used to screen these two lines. All the lines which are in the pedigree of 'Forno', 'Forel' and 'Fiorina' will also be screened with the closest *Lr75* flanking markers to elucidate the origin of *Lr75*. To determine the presence of *Lr75* in different wheat lines, we will screen 350 additional winter, spring and spelt wheat cultivars from different regions of the world with the *Lr75* closest flanking markers. Out of these 350 lines, 327 lines belong to the European origin, 19 lines originate from North American origin, three from South America and one from Indian origin. All the positive lines that contain *Lr75* will later be screened for leaf rust resistance response at seedling as well as adult plant stage to test the effect of *Lr75* in these lines. Information on the origin and presence of *Lr75* in different wheat lines will help to to effectively use this gene in wheat breeding programmes.

5.3 High-throughput genotyping in wheat

In the past few years, high throughput genotyping in wheat was considered as a tedious and challenging task because of the lack of availability of a reference genome sequence of wheat. However, a major advancement has been done by the scientific community in the past few years to sequence the reference genomes of crops of agronomical importance such as rice, sorghum, maize and barley (International Rice Genome Sequencing Project 2005; Paterson et al. 2009; Schnable et al. 2009; Mayer et al. 2012). Significant progress has also been made to establish the reference draft genome sequences of progenitor species of A and D genome donor of bread wheat (Ling et al. 2013; Jia et al. 2013). Recently, the efforts of International Wheat Genome Sequencing Consortium (IWGSC) led to the establishment of a draft genome sequence of hexaploid wheat cv. 'Chinese Spring' (The International Wheat Genome Sequencing Consortium 2014). Though the reference draft genomes of diploid wheat genome donors and hexaploid wheat cv. 'Chinese Spring' are available, there is a need for whole genome sequencing of several wheat haplotypes/ genotypes and this is still

an expensive and difficult task. Therefore, alternative strategies are needed. Targeted-capture resequencing of part of the wheat genome such as exome capture are highly useful.

5.3.1 Exome capture as an efficient tool for targeted-capture resequencing of parts of the wheat genome

Exome capture involves the sequencing of the transcribed regions of the genome which comprises 1-2% of the whole genome (Paux et al. 2006; Warr et al. 2015). Exome capture is based on the use of solid- or liquid-phase hybridization assays for hybridizing the sequence of interest by using long oligonucleotide probes as baits (Albert et al. 2007; Teer and Mullikin 2010; Saintenac et al. 2011; Winfield et al. 2012). Exome capture kits have been designed for many crops such as maize, soybean, cotton, barley and tetraploid wheat (Fu et al. 2010; Haun et al. 2011; Saintenac et al. 2011; Salmon et al. 2012; Mascher et al. 2014). In maize, Fu et al. (2010) described the use of NimbleGen array-based sequence capture and identified 2500 high-quality SNPs between the target sequences of the reference maize genomes Mo17 and B73 in a 2.2 Mb region. Saintenac et al. (2011) used liquid-phase targeted-capture approach for enrichment of genomic libraries with sequences of interest in tetraploid wheat. Winfield et al. (2012) designed an exome capture kit for hexaploid wheat based on 100,000 transcripts collected from eight UK wheat cultivars. The exome capture kits have also been designed for identifying induced mutations in the genome in order to investigate gene function in rice and tetraploid wheat (Henry et al. 2014). Exome capture kits have also been developed for identification of new SNP-based markers in the wheat genome (Allen et al. 2013). Allen et al. (2013) generated 95,266 putative SNPs from the sequences of wheat cultivars generated by targeted resequencing of the wheat exome.

As compared to exome sequencing which only covers the coding regions of the genome, the technique for SNP development used in our study also covers the non-coding regions of the genome. Due to the low conservation of these regions among three sub-genomes of wheat and among wheat cultivars, the chances of getting polymorphic SNPs are higher. Exome capture is an efficient tool for sequencing the coding regions with high coverage, but studies have shown that it fails to capture the non-coding regions as well as regulatory mutations which are present in promoter and enhancer regions (Schneeberger 2014; Warr et al. 2015). However, to increase the utility of exome capture, studies have shown that it is possible to extend the exome capture regions to capture the non-coding regions and regulatory elements that include promoters and enhancers (Samuels et al. 2013). This will increase the efficiency of exome capture to sequence all the functional regions of the genome. Therefore, re-sequencing of the reduced representations of the wheat genome serves as an efficient tool in SNP identification and genetic mapping in wheat.

5.3.2 SNP genotyping in wheat

Single nucleotide polymorphisms represent the most frequent source of genetic variation in the wheat genome and they are commonly used for diversity analysis in wheat (Akhunov et al. 2009; van Poecke et al. 2013). In the past few years, a number of SNP genotyping arrays have been designed in wheat such as Illumina GoldenGate assay which can be used for genotyping SNPs in homozygous tetraploid and hexaploid wheat lines (Akhunov et al. 2009). Allen et al. (2011) used KASPar assays for genotyping of SNPs which were designed from the NCBI wheat EST sequences and NGS cDNA sequences of UK wheat cultivars. Cavanagh et al. (2013) by using the transcriptomic data from different wheat accessions from Australia and USA developed a 9K Infinium wheat SNP assay. Using this assay, they positioned 7,504 SNPs on a wheat consensus SNP map comprised of seven different mapping populations. Li et al. (2013) also used this assay for the identification of a novel wheat gene, *H34* which confers resistance to Hessian fly (*Mayetiola destructor*). Additionally, the wheat 90K SNP array comprising 90,000 gene-associated SNPs was recently developed by Wang et al. (2014) for characterization of genetic variation in tetraploid and hexaploid wheat populations. All the so far developed SNP arrays provide a useful resource for genome-wide studies including large number of individuals or mapping populations in wheat.

However, due to the presence of repetitive regions, homeologous and paralogous copies, SNP mapping in polyploid wheat is a challenging task. This can lead to polymorphism detection between homeologous sequences or non-amplification instead of SNP detection between the regions of interest. To overcome these problems, the possible solutions could be 1) to develop the new array for genotyping the wheat populations which can verify homeologous-specific SNP markers or 2) design homeologous-specific SNP markers from the regions that are less conserved as compared to genes. The most appropriate way to get homeologous-specific SNP markers is to isolate the chromosome of interest by using flow cytometric chromosome sorting. Vrána et al. (2015) by using this technique have successfully isolated 15 out of 21 chromosomes without contamination of the homeologous chromosomes. The purity of the isolated chromosomes can range from 16% to 80%. We have successfully used the sequenced flow-sorted chromosome 1B of wheat cultivars ‘Arina’ and ‘Forno’ for SNP marker development which resulted in 67% of genome-specific polymorphic SNP markers in the RIL population. In addition, if the mapping population is heterozygous and derived from parents that have a very similar genetic background, then chances of detecting polymorphism in the mapping population using the available SNP arrays are very low. We have used in our study the wheat 90K SNP array developed by Wang et al. (2014) in order to test the efficiency of the array on a BC_3F_2 heterozygous population. The analysis revealed that 35% SNPs detected less than 50% heterozygous alleles. We conclude that the 90K SNP array might not be able to reliably detect heterozygous alleles. Therefore, it will be more feasible to develop genotype-specific SNP assays by detecting gene-associated SNPs between the cultivars of interest using syntenic information from the model grass

genomes such as barley, rice and *Brachypodium*. However, advances in sequencing of diploid A and D wheat genome progenitors as well as a draft genome of hexaploid wheat cv. 'Chinse Spring' will probably increase the efficiency of genotype-specific SNP mapping in wheat (Ling et al. 2013; Jia et al. 2013; The International Wheat Genome Sequencing Consortium 2014).

5.4 Cloning of *Lr75* and approaches for validation of candidate genes

High resolution genetic and physical mapping of *Lr75* in a backcross derived NIL population mapped this gene to a 0.96 cM interval which corresponds to a 723 kb physical region in 'Chinese Spring' and a yet incomplete 403.6 kb physical region with one gap in *Lr75* donor cv. 'Forno'. In order to complete the molecular isolation of *Lr75*, we will establish the complete physical region of the donor cv. 'Forno' and identify the candidate genes between the closest genetic flanking markers by comparing the model with the genes of *Brachypodium* and other well-studied grass genomes. We generated a set of EMS mutants in a NIL of 'Arina' with *Lr75* and *QLr.sfr-7BL* and identified more than ten *Lr75* loss-of-function mutants. We will test by genetic complementation analysis if the mutations are indeed in *Lr75* and will sequence the putative candidate genes in the mutants and identify the mutations in the coding regions of the candidate genes. Once the candidate gene is identified we will functionally validate it by using stable transformation assay. Further functional analysis will depend on the type of gene isolated. The presence in the target interval in 'Chinese Spring' and 'Forno' of cyclin like *F-box* gene is very promising. In addition, to *F-box* gene, two copies of a chimeric gene encoding a domain of unknown function (DUF) with a serine/threonine kinase domain are also interesting candidates.

References

Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. Biophotonics Int 11:36–42.

Afzal AJ, Wood AJ, Lightfoot DA (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol Plant Microbe Interact 21:507–517. doi: 10.1094/MPMI-21-5-0507

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27:617–631. doi: 10.1007/s00299-008-0507-z

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet 119:507–517. doi: 10.1007/s00122-009-1059-5

Akhunov ED, Akhunova a. R, Dvořák J (2005) BAC libraries of *Triticum urartu*, *Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. Theor Appl Genet 111:1617–1622. doi: 10.1007/s00122-005-0093-1

Akhunov ED, Akhunova AR, Anderson OD, et al (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11:702. doi: 10.1186/1471-2164-11-702

Akhunov ED, Akhunova AR, Linkiewicz AM, et al (2003) Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. Proc Natl Acad Sci U S A 100:10836–10841. doi: 10.1073/pnas.1934431100

Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA (2007) Direct selection of human genomic loci by microarray hybridization. Nat Methods 4:903–905. doi: 10.1038/nmeth1111

Alexandratos N, Bruinsma J (2012) World Agriculture Towards 2030/2050: The 2012 Revision. ESA Working paper No. 12-03. FAO, Rome.

Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum *aestivum* L.). Plant Biotechnol J 1–14. doi: 10.1111/j.1467-7652.2011.00628.x

Allen AM, Barker GLA, Wilkinson P, Burridge A, Winfield M, Coghill J, Uauy C, Griffiths S, Jack P, Berry S, Werner P, Melichar JPE, McDougall J, Gwilliam R, Robinson P, Edwards KJ (2013) Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). Plant Biotechnol J 11:279–295. doi: 10.1111/pbi.12009

Bai D, Knott DR (1992) Suppression of rust resistance in bread wheat (*Triticum aestivum* L.) by D-genome chromosomes. Genome 35:276–282.

Banks L, Pim D, Thomas M (2003) Viruses and the 26S proteasome: hacking into destruction. Trends Biochem Sci 28:452–459. doi: 10.1016/S0968-0004(03)00141-5

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51:910–918. doi: 10.1111/j.1365-313X.2007.03193.x

Bariana H, Miah H, Brown G, Willey N, Lehmensiek A (2007) Molecular mapping of durable rust resistance in wheat and its implication in breeding. In: Wheat production in stressed

environments: Proceedings of the 7th international wheat Conf, Mar del Plata, Argentina, 27 November–2 December 2005 pp 723-728.

Barker GLA, Edwards KJ (2009) A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. Plant Biotechnol J 7:318–325. doi: 10.1111/j.1467-7652.2009.00412.x

Bernardo AN, Bradbury PJ, Ma H, Hu S, Bowden RL, Buckler ES, Bai G (2009) Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. BMC Genomics 10:251. doi: 10.1186/1471-2164-10-251

Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo M, Salvi S, Jansen J, Viola R, Gut I, Laurens F, Chagné D, Velasco R, van de Weg E, Troggio M (2014) Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (*Malus* × *domestica* Borkh). PLoS One 9:e110377. doi: 10.1371/journal.pone.0110377

Bolton MD, Kolmer JA, Garvin DF (2008) Wheat leaf rust caused by *Puccinia triticina*. Mol Plant Pathol 9:563–575. doi: 10.1111/j.1364-3703.2008.00487.x

Breen J, Wicker T, Shatalina M, et al (2013) A physical map of the short arm of wheat chromosome 1A. PLoS One. doi: 10.1371/journal.pone.0080272

Brenchley R, Spannagl M, Pfeifer M, et al (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705–10. doi: 10.1038/nature11650

Brown JKM, Hovmøller MS (2002) Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. Science 297:537–541. doi: 10.1126/science.1072678

Brueggeman R, Druka A, Nirmala J, Cavileer T, Drader T, Rostoks N, Mirlohi A, Bennypaul H, Gill U, Kudrna D, Whitelaw C, Kilian A, Han F, Sun Y, Gill K, Steffenson B, Kleinhofs A (2008) The stem rust resistance gene *Rpg5* encodes a protein with nucleotide-binding-site, leucine-rich, and protein kinase domains. Proc Natl Acad Sci U S A 105:14970–14975. doi: 10.1073/pnas.0807270105

Buckler ES, Holland JB, Bradbury PJ, et al (2009) The genetic architecture of maize flowering time. Science 325:714–718. doi: 10.1126/science.1174276

Buerstmayr M, Matiasch L, Mascher F, Vida G, Ittu M, Robert O, Holdgate S, Flath K, Neumayer A, Buerstmayr H (2014) Mapping of quantitative adult plant field resistance to leaf rust and stripe rust in two European winter wheat populations reveals co-location of three QTL conferring resistance to both rust pathogens. Theor Appl Genet 127:2011–2028. doi: 10.1007/s00122-014-2357-0

Caldwell R (1968) Breeding for general and/or specific plant disease resistance. In: K.W.Finley and K.W. Shepherd (ed) Proceeding International Wheat Genetic Symposium 3rd, Canberra, pp 263-272.

Cavanagh CR, Chao S, Wang S, et al (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci U S A 110:8057–62. doi: 10.1073/pnas.1217133110

Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo MC, Dubcovsky J (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. Mol Breed 23:23–33. doi: 10.1007/s11032-008-9210-6

Charmet G (2011) Wheat domestication: Lessons for the future. C R Biol 334:212–220. doi: 10.1016/j.crvi.2010.12.013

Chelaifa H, Chagué V, Chalabi S, Mestiri I, Arnaud D, Deffains D, Lu Y, Belcram H, Huteau V,

Chiquet J, Coriton O, Just J, Jahier J, Chalhoub B (2013) Prevalence of gene expression additivity in genetically stable wheat allohexaploids. New Phytol 197:730–736. doi: 10.1111/nph.12108

Chen W, Liu T, Gao L (2013) Suppression of stripe rust and leaf rust resistances in interspecific crosses of wheat. Euphytica 192:339–346. doi: 10.1007/s10681-012-0854-2

Chen XM (2005) Review / Synthèse Epidemiology and control of stripe rust [*Puccinia striiformis* f . sp . *tritici*] on wheat. Can J Bot 27:314–337. doi: 10.1071/ar07045

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19. doi: 10.1186/1471-2156-3-19

Choulet F, Wicker T, Rustenholz C, et al (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701. doi: 10.1105/tpc.110.074187

Cloutier S, McCallum BD, Loutre C, Banks TW, Wicker T, Feuillet C, Keller B, Jordan MC (2007) Leaf rust resistance gene *Lr1*, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large psr567 gene family. Plant Mol Biol 65:93–106. doi: 10.1007/s11103-007-9201-8

Das M, Rajaram S, Kronstad W, Mundt C, Singh R (1993) Associations and genetics of three components of slow rusting in leaf rust of wheat. Euphytica 68:99–109.

Devos KM (2005) Updating the "crop circle." Curr Opin Plant Biol 8:155–162. doi: 10.1016/j.pbi.2005.01.005

Dixon J, Braun HJ, Crouch J (2009) Transitioning wheat research to serve the future needs of the developing world. In: Dixon J, Braun HJ, Kosina P, editors. Wheat facts and futures. Mexico: CIMMYT. 1–19.

Doležel J, Kubaláková M, Paux E, Bartoš J, Feuillet C (2007) Chromosome-based genomics in the cereals. Chromosom Res 15:51–66. doi: 10.1007/s10577-006-1106-x

Doležel J, Vrána J, Safář J, Bartoš J, Kubaláková M, Simková H (2012) Chromosomes in the flow to simplify genome analysis. Funct Integr Genomics 12:397–416. doi: 10.1007/s10142-012-0293-0

Dreher K, Callis J (2007) Ubiquitin, hormones and biotic stress in plants. Ann Bot 99:787–822. doi: 10.1093/aob/mcl255

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316:1862–1866. doi: 10.1126/science.1143986

Duplessis S, Cuomo CA, Lin Y-C, et al (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci 108:9166–9171. doi: 10.1073/pnas.1019315108

Dyck P, Samborski D (1979) Adult-plant leaf rust resistance in PI 250413, an introduction of common wheat. Can J Plant Sci 59:329–332.

Eckardt NA (2008) Grass Genome Evolution. Plant Cell 20:3–4. doi: 10.1105/tpc.108.058586

Endo TR, Gill BS (1996) The Deletion Stocks of Common Wheat. J Hered 87:295–307. doi: 10.1093/oxfordjournals.jhered.a023003

FAOstat (2014) Food and agricultural organization of the United Nations, FAOstat database.

Faris JD, Liu Z, Xu SS (2013) Genetics of tan spot resistance in wheat. Theor Appl Genet 126:2197–2217. doi: 10.1007/s00122-013-2157-y

Faris JD, Zhang Z, Lu H, Lu S, Reddy L, Cloutier S, Fellers JP, Meinhardt SW, Rasmussen JB, Xu SS, Oliver RP, Simons KJ, Friesen TL (2010) A unique wheat disease resistance-like gene

governs effector-triggered susceptibility to necrotrophic pathogens. Proc Natl Acad Sci U S A 107:13544–13549. doi: 10.1073/pnas.1004090107

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci U S A 100:15253–15258. doi: 10.1073/pnas.2435133100

Flavell RB, Rimpau J, Smith DB (1977) Repeated sequence DNA relationships in four cereal genomes. Chromosoma 63:205–222.

Flor HH (1955) Host-parasite interactions in flax rust: its genetics and other implications. Phytopathology 45:680–685.

Frenkel Z, Paux E, Mester D, Feuillet C, Korol A (2010) LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. BMC Bioinformatics 11:584. doi: 10.1186/1471-2105-11-584

Fu D, Uauy C, Distelfeld A, Blechl A, Epstein L, Chen X, Sela H, Fahima T, Dubcovsky J (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323:1357–1360. doi: 10.1126/science.1166289

Fu Y, Springer NM, Gerhardt DJ, Ying K, Yeh CT, Wu W, Swanson-Wagner R, D'Ascenzo M, Millard T, Freeberg L, Aoyama N, Kitzman J, Burgess D, Richmond T, Albert TJ, Barbazuk WB, Jeddeloh JA, Schnable PS (2010) Repeat subtraction-mediated sequence capture from a complex genome. Plant J 62:898–909. doi: 10.1111/j.1365-313X.2010.04196.x

Gale M, Atkinson M, Chinoy C, Harcourt R, Jia J, Li Q, Devos K (1995) Genetic maps of hexaploid wheat. In: Li ZS, Xin ZY (eds) Proc 8th Int Wheat Genet Symp, China Agricultural Scientech Press, Beeijing, China. 29–40.

Gómez-Gómez L, Boller T (2000) *FLS2*: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol Cell 5:1003–11.

Goyeau H, Ammar K, Berder J (2010) Virulence in *Puccinia triticina* for durum wheat cultivar Creso and other durum wheat cultivars carrying resistance gene *Lr14a* in France. Plant Dis 94:1068. doi: 10.1094/PDIS-94-8-1068A

Grattapaglia D, Silva-Junior OB, Kirst M, de Lima BM, Faria DA, Pappas GJ (2011) High-throughput SNP genotyping in the highly heterozygous genome of Eucalyptus: assay success, polymorphism and transferability across species. BMC Plant Biol 11:65. doi: 10.1186/1471-2229-11-65

GRIS (2014) GRIS. Genetic Resources Information System for Wheat and Triticale.

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP, Stupar RM (2011) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol 155:645–655. doi: 10.1104/pp.110.166736

Henry IM, Nagalakshmi U, Lieberman MC, Ngo KJ, Krasileva K V., Vasquez-Gross H, Akhunova A, Akhunov E, Dubcovsky J, Tai TH, Comai L (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell 26:1382–1397. doi: 10.1105/tpc.113.121590

Hernandez P, Martis M, Dorado G, Pfeifer M, Gálvez S, Schaaf S, Jouve N, Šimková H, Valárik M, Doležel J, Mayer KFX (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. Plant J 69:377–386. doi: 10.1111/j.1365-313X.2011.04808.x

Herrera-Foessel SA, Lagudah ES, Huerta-Espino J, Hayden MJ, Bariana HS, Singh D, Singh RP

(2011) New slow-rusting leaf rust and stripe rust resistance genes *Lr67* and *Yr46* in wheat are pleiotropic or closely linked. Theor Appl Genet 122:239–249. doi: 10.1007/s00122-010-1439-x

Herrera-Foessel SA, Singh RP, Huerta-Espino J, Crossa J, Djurle A, Yuen J (2008) Genetic analysis of slow rusting resistance to leaf rust in durum wheat. Crop Sci 48:2132–2140.

Herrera-Foessel SA, Singh RP, Huerta-Espino J, Rosewarne GM, Periyannan SK, Viccars L, Calvo-Salazar V, Lan C, Lagudah ES (2012) *Lr68*: a new gene conferring slow rusting resistance to leaf rust in wheat. Theor Appl Genet 124:1475–1486. doi: 10.1007/s00122-012-1802-1

Hiebert CW, Thomas JB, Somers DJ, McCallum BD, Fox SL (2007) Microsatellite mapping of adult-plant leaf rust resistance gene *Lr22a* in wheat. Theor Appl Genet 115:877–884. doi: 10.1007/s00122-007-0604-3

Hiebert CW, Thomas JB, McCallum BD, Humphreys DG, DePauw RM, Hayden MJ, Mago R, Schnippenkoetter W, Spielmeyer W. (2010) An introgression on wheat chromosome 4DL in RL6077 (Thatcher*6/PI 250413) confers adult plant resistance to stripe rust and leaf rust (*Lr67*). Theor Appl Genet 121:1083–1091. doi: DOI:10.1007/s00122-010-1373-y

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419–1428. doi: 10.1101/gr.091678.109

Huang L, Brooks S a, Li W, Fellers JP, Trick HN, Gill BS (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164:655–664.

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid Acetyl-CoA carboxylase and 3-Phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci U S A 99:8133–8138.

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. 1068–1076. doi: 10.1101/gr.089516.108.

Huerta-Espino J, Singh RP, Germán S, McCallum BD, Park RF, Chen WQ, Bhardwaj SC, Goyeau H (2011) Global status of wheat leaf rust caused by *Puccinia triticina*. Euphytica 179:143–160. doi: 10.1007/s10681-011-0361-x

Hurni S, Brunner S, Buchmann G, Herren G, Jordan T, Krukowski P, Wicker T, Yahiaoui N, Mago R, Keller B (2013) Rye *Pm8* and wheat *Pm3* are orthologous genes and show evolutionary conservation of resistance function against powdery mildew. Plant J 76:957–969. doi: 10.1111/tpj.12345

Hurni S, Brunner S, Stirnweis D, Herren G, Peditto D, Mcintosh RA, Keller B (2014) The powdery mildew resistance gene *Pm8* derived from rye is suppressed by its wheat ortholog *Pm3*. Plant J 1–10. doi: 10.1111/tpj.12593

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800. doi: 10.1038/nature03895

Janda J, Bartos J, Safár J, Kubaláková M, Valárik M, Číhalíková J, Simková H, Caboche M, Sourdille P, Bernard M, Chalhoub B, Dolezel J (2004) Construction of a subgenomic BAC library specific for chromosomes 1D, 4D and 6D of hexaploid wheat. Theor Appl Genet 109:1337–1345. doi: 10.1007/s00122-004-1768-8

Jia J, Zhao S, Kong X, et al (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95. doi: 10.1038/nature12028

Jin Y (2011) Role of *Berberis* spp. as alternate hosts in generating new races of *Puccinia graminis* and P. *striiformis*. Euphytica 179:105–108. doi: 10.1007/s10681-010-0328-3

Jin Y, Szabo LJ, Carson M (2010) Century-old mystery of *Puccinia striiformis* life history solved with the identification of *Berberis* as an alternate host. Phytopathology 100:432–435. doi: 10.1094/PHYTO-100-5-0432

Johnson R (1981) Durable resistance: Definition of, genetic control, attainment in plant breeding. Am Phytopathol Soc 71:567–568.

Jonah PM, Bello LL, Lucky O, Midau A, Moruppa SM (2011) Review : The importance of molecular markers in plant breeding programmes.

Kema GHJ, Lange W, Van Silfhout CH (1995) Differential suppression of stripe rust resistance in synthetic wheat hexaploids derived from *Triticum turgidum* subsp. *dicoccoides* and *Aegilops squarrosa*. Phytopathology 85:425–429.

Kerber ER, Green GJ (1980) Suppression of stem rust resistance in hexaploid wheat cv. Canthatch by chromosome 7DL. Can J Bot 58:1347–1350.

Knott DR (2000) Inheritance of resistance to stem rust in Medea durum-wheat and the role of suppressors. Crop Sci 40:98–102.

Kolmer JA (2013) Leaf Rust of Wheat: Pathogen Biology, Variation and Host Resistance. Forests 4:70–84. doi: 10.3390/f4010070

Kolmer JA (2005) Tracking wheat rust on a continental scale. Curr Opin Plant Biol 8:441–449. doi: 10.1016/j.pbi.2005.05.001

Kolmer JA, Singh RP, Garvin DF, Viccars L, William HM, Huerta-Espino J, Ogbonnaya FC, Raman H, Orford S, Bariana HS, Lagudah ES (2008) Analysis of the Rust Resistance Region in Wheat Germplasm. Crop Sci 48:1841–1852. doi: 10.2135/cropsci2007.08.0474

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175.

Kota R, Spielmeyer W, McIntosh RA, Lagudah ES (2006) Fine genetic mapping fails to dissociate durable stem rust resistance gene *Sr2* from pseudo-black chaff in common wheat (*Triticum aestivum* L.). Theor Appl Genet 112:492–499. doi: 10.1007/s00122-005-0151-8

Krattinger S, Wicker T, Keller B (2007) Map-Based Cloning of Genes in Triticeae (Wheat and Barley). 1–21.

Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–1363. doi: 10.1126/science.1166453

Kubaláková M, Vrána J, Číhalíková J, Simková H, Dolezel J (2002) Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 104:1362–1372. doi: 10.1007/s00122-002-0888-2

Lagudah ES, Krattinger SG, Herrera-Foessel S, Singh RP, Huerta-Espino J, Spielmeyer W, Brown-Guedira G, Selter LL, Keller B (2009) Gene-specific markers for the wheat gene *Lr34/Yr18/Pm38* which confers resistance to multiple fungal pathogens. Theor Appl Genet 119:889–898. doi: 10.1007/s00122-009-1097-z

Lagudah ES, McFadden H, Singh RP, Huerta-Espino J, Bariana HS, Spielmeyer W (2006) Molecular genetic characterization of the *Lr34/Yr18* slow rusting resistance gene region in wheat. Theor Appl Genet 114:21–30. doi: 10.1007/s00122-006-0406-z

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (1987)

MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181. doi: 10.1016/j.ygeno.2008.12.003

Lechner E, Achard P, Vansiri A, Potuschak T, Genschik P (2006) F-box proteins everywhere. Curr Opin Plant Biol 9:631–638. doi: 10.1016/j.pbi.2006.09.003

Li C, Chen M, Chao S, Yu J, Bai G (2013) Identification of a novel gene, *H34*, in wheat using recombinant inbred lines and single nucleotide polymorphism markers. Theor Appl Genet 126:2065–2071. doi: 10.1007/s00122-013-2118-5

Lillemo M, Asalf B, Singh RP, Huerta-Espino J, Chen XM, He ZH, Bjørnstad Å (2008) The adult plant rust resistance loci *Lr34/Yr18* and *Lr46/Yr29* are important determinants of partial resistance to powdery mildew in bread wheat line Saar. Theor Appl Genet 116:1155–1166. doi: 10.1007/s00122-008-0743-1

Lillemo M, Singh R, William M, Herrera-Foessel S, Huerta- Espino J, German S, Campos P, Chaves M, Madriaga R, Xia X, Liang S, Liu D, Li Z, Lagudah E (2011) Multiple rust resistance and gene additivity in wheat: lessons from multi-location case studies in the cultivars Parula and Saar. Global Rust Initiative Meeting, St. Paul. 111–120.

Lindhout P (2002) The perspectives of polygenic resistance in breeding for durable disease resistance. Euphytica 124:217–226. doi: 10.1023/A:1015686601404

Ling H-Q, Zhao S, Liu D, et al (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87–90. doi: 10.1038/nature11997

Liu W, Frick M, Huel R, Nykiforuk CL, Wang X, Gaudet DA, Eudes F, Conner RL, Kuzyk A, Chen Q, Kang Z, Laroche A (2014) The stripe rust resistance gene *Yr10* encodes an evolutionary-conserved and unique CC–NBS–LRR sequence in wheat. Mol Plant 7:1740–1755. doi: 10.1093/mp/ssu112

Lucas SJ, Akpınar BA, Kantar M, Weinstein Z, Aydınoğlu F, Safář J, Simková H, Frenkel Z, Korol A, Magni F, Cattonaro F, Vautrin S, Bellec A, Bergès H, Doležel J, Budak H (2013) Physical mapping integrated with syntenic analysis to characterize the gene space of the long arm of wheat chromosome 1A. PLoS One 8:e59542. doi: 10.1371/journal.pone.0059542

Lucas SJ, Hana Š, Jan Š, et al (2012) Functional features of a single chromosome arm in wheat (1AL) determined from its structure. Funct Integr Genomics 12:173–182. doi: 10.1007/s10142-011-0250-3

Maccaferri M, Mantovani P, Tuberosa R, Deambrogio E, Giuliani S, Demontis A, Massi A, Sanguineti MC (2008) A major QTL for durable leaf rust resistance widely exploited in durum wheat breeding programs maps on the distal region of chromosome arm 7BL. Theor Appl Genet 117:1225–1240. doi: 10.1007/s00122-008-0857-5

Maccaferri M, Zhang J, Bulli P, Abate Z, Chao S, Cantu D, Bossolini E, Chen X, Pumphrey M, Dubcovsky J (2015) A Genome-Wide Association Study of Resistance to Stripe Rust (*Puccinia striiformis* f. sp. *tritici*) in a Worldwide Collection of Hexaploid Spring Wheat (*Triticum aestivum* L.). G3: Genes|Genomes|Genetics 5:449–465. doi: 10.1534/g3.114.014563

Mago R, Tabe L, McIntosh R a., Pretorius Z, Kota R, Paux E, Wicker T, Breen J, Lagudah ES, Ellis JG, Spielmeyer W (2011) A multiple resistance locus on chromosome arm 3BS in wheat confers resistance to stem rust (*Sr2*), leaf rust (*Lr27*) and powdery mildew. Theor Appl Genet 123:615–623. doi: 10.1007/s00122-011-1611-y

Mago R, Tabe L, Vautrin S, Šimková H, Kubaláková M, Upadhyaya N, Berges H, Kong X, Breen J, Doležel J, Appels R, Ellis JG, Spielmeyer W (2014) Major haplotype divergence including multiple germin-like protein genes, at the wheat *Sr2* adult plant stem rust resistance locus. BMC

Plant Biol 14:1–11. doi: 10.1186/s12870-014-0379-z

Mascher M, Jost M, Kuon J-E, Himmelbach A, Aßfalg A, Beier S, Scholz U, Graner A, Stein N (2014) Mapping-by-sequencing accelerates forward genetics in barley. Genome Biol 15:R78. doi: 10.1186/gb-2014-15-6-r78

Mayer KFX, Martis M, Hedley PE, et al (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263. doi: 10.1105/tpc.110.082537

Mayer KFX, Waugh R, Brown JWS, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ, Wise RP, Stein N (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716. doi: 10.1038/nature11543

Mcintosh RA, Dubcovsky J, Rogers JM, Morris C, Appels R, Xia X (2014) Catalogue of gene symbols for wheat: 2013-2014.

McIntosh RA, Dubcovsky J, Rogers WJ, Morris C, Appels R, Xia XC (2012) Catalogue of gene symbols for wheat: 2012. In KOMUGI Integrated Wheat Science Database.

McIntosh RA, Dyck PL (1975)) Cytogenetical studies in wheat VII. Gene *Lr23* for reaction to *Puccinia recondita* in Gabo and related cultivars. Aust J Biol Sci 28:201–211.

Mcintosh RA, Wellings CR, Park RF (1995) Wheat rusts: An atlas of resistance genes. Dordrecht: Kluwer Academic Publishers.

Mcintosh RA, Zhang P, Cowger C, Parks R, Lagudah ES, Hoxha S (2011) Rye-derived powdery mildew resistance gene *Pm8* in wheat is suppressed by the *Pm3* locus. Theor Appl Genet 123:359–367. doi: 10.1007/s00122-011-1589-5

Mendgen K, Hahn M (2002) Plant infection and the establishment of fungal biotrophy. Trends Plant Sci 7:352–356. doi: 10.1016/S1360-1385(02)02297-5

Messmer MM, Seyfarth R, Keller M, Schachermayr G, Winzeler M, Zanetti S, Feuillet C, Keller B (2000) Genetic analysis of durable leaf rust resistance in winter wheat. Theor Appl Genet 100:419–431.

Mochida K, Yoshida T, Sakurai T, Ogihara Y, Shinozaki K (2009) TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. Plant Physiol 150:1135–1146. doi: 10.1104/pp.109.138214

Molnár I, Kubaláková M, Šimková H, Farkas A, Cseh A, Megyeri M, Vrána J, Molnár-Láng M, Doležel J (2014) Flow cytometric chromosome sorting from diploid progenitors of bread wheat, *T. urartu*, *Ae. speltoides* and *Ae. tauschii*. Theor Appl Genet 1–14. doi: 10.1007/s00122-014-2282-2

Moore JW, Herrera-Foessel S, Lan C, Schnippenkoetter W, Ayliffe M, Huerta-Espino J, Lillemo M, Viccars L, Milne R, Periyannan S, Kong X, Spielmeyer W, Talbot M, Bariana H, Patrick JW, Dodds P, Singh R, Lagudah E (2015) A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nat Genet 47:1494–1498. doi: 10.1038/ng.3439

Nelson JC, Singh RP, Autrique JE, Sorrells ME (1997) Mapping genes conferring and suppressing leaf rust resistance in wheat. Crop Sci 37:1928–1935.

Oerke EC, Dehne HW, Schoenbeck F, Weber A (1994) Crop production and crop protection: Estimated losses in major food and cash crops.Elsevier Science Publishers, Amsterdam.

Paillard S, Schnurbusch T, Winzeler M, Messmer M, Sourdille P, Abderhalden O, Keller B, Schachermayr G (2003) An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.). Theor Appl Genet 107:1235–1242. doi: 10.1007/s00122-003-1361-6

Park RF, Goyeau H, Felsenstein FG, Bartoš P, Zeller FJ (2001) Regional phenotypic diversity of

Puccinia triticina and wheat host resistance in western Europe, 1995. Euphytica 122:113–127. doi: 10.1023/A:1012603500686

Paterson AH, Bowers JE, Bruggmann R, et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556. doi: 10.1038/nature07723

Pathan AK, Park RF (2006) Evaluation of seedling and adult plant resistance to leaf rust in European wheat cultivars. Euphytica 149:327–342. doi: 10.1007/s10681-005-9081-4

Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant J-P, Sourdille P, Balfourier F, Le Paslier M-C, Chauveau A, Cakir M, Gandon B, Feuillet C (2010) Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol J 8:196–210. doi: 10.1111/j.1467-7652.2009.00477.x

Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J 48:463–74. doi: 10.1111/j.1365-313X.2006.02891.x

Paux E, Sourdille P, Mackay I, Feuillet C (2012) Sequence-based marker development in wheat: Advances and applications to breeding. Biotechnol Adv 30:1071–1088. doi: 10.1016/j.biotechadv.2011.09.015

Paux E, Sourdille P, Salse J, et al (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101–104. doi: 10.1126/science.1161847

Pazhouhandeh M, Dieterle M, Marrocco K, Lechner E, Berry B, Brault V, Hemmer O, Kretsch T, Richards KE, Genschik P, Ziegler-Graff V (2006) F-box-like domain in the polerovirus protein P0 is required for silencing suppressor function. Proc Natl Acad Sci USA 103:1994–1999. doi: 10.1073/pnas.0510784103

Peng JH, Sun D, Nevo E (2011) Domestication evolution, genetics and genomics in wheat. Mol Breed 28:281–301. doi: 10.1007/s11032-011-9608-4

Periyannan S, Moore J, Ayliffe M, Bansal U, Wang X, Huang L, Deal K, Luo M, Kong X, Bariana H, Mago R, McIntosh RA, Dodds P, Dvorak J, Lagudah E (2013) The Gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. Science (80-) 341:786–789.

Pink DAC (2002) Strategies using genes for non-durable disease resistance. Euphytica 124:227–236.

Raats D, Frenkel Z, Krugman T, et al (2013) The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. Genome Biol 14:R138. doi: 10.1186/gb-2013-14-12-r138

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, Beckert M, Charmet G (2006) Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome 49:1131–9. doi: 10.1139/g06-067

Ribeiro Do Vale FX, Parlevliet JE, Zambolim L (2001) Concepts in plant disease resistance. Fitopatol Bras 26:577–589.

Roelfs AP, Singh RP, Saari EE (1992) Rust diseases of wheat: concepts and methods of disease management. CIMMYT, Mexico, D.F.

Rosewarne GM, Singh RP, Huerta-Espino J, Herrera-Foessel SA, Forrest KL, Hayden MJ, Rebetzke GJ (2012) Analysis of leaf and stripe rust severities reveals pathotype changes and multiple minor QTLs associated with resistance in an Avocet × Pastor wheat population. Theor Appl Genet 124:1283–1294. doi: 10.1007/s00122-012-1786-x

Rosewarne GM, Singh RP, Huerta-Espino J, William HM, Bouchet S, Cloutier S, McFadden H, Lagudah ES (2006) Leaf tip necrosis, molecular markers and beta1-proteasome subunits

associated with the slow rusting resistance genes *Lr46/Yr29*. Theor Appl Genet 112:500–8. doi: 10.1007/s00122-005-0153-6

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R (2005) Single-feature polymorphism discovery in the barley transcriptome. Genome Biol 6:R54. doi: 10.1186/gb-2005-6-6-r54

Šafář J, Bartoš J, Janda J, Bellec A, Kubaláková M, Valárik M, Pateyron S, Weiserová J, Tušková R, Číhalíková J, Vrána J, Šimková H, Faivre-Rampant P, Sourdille P, Caboche M, Bernard M, Doležel J, Chalhoub B (2004) Dissecting large and complex genomes: Flow sorting and BAC cloning of individual chromosomes from bread wheat. Plant J 39:960–968. doi: 10.1111/j.1365-313X.2004.02179.x

Safár J, Simková H, Kubaláková M, Číhalíková J, Suchánková P, Bartos J, Doležel J (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenet Genome Res 129:211–223. doi: 10.1159/000313072

Saintenac C, Jiang D, Akhunov ED (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol 12:R88. doi: 10.1186/gb-2011-12-9-r88

Saintenac C, Jiang D, Wang S, Akhunov E (2013a) Sequence-based mapping of the polyploid wheat genome. G3Genes|Genomes|Genetics 3:1105–1114. doi: 10.1534/g3.113.005819

Saintenac C, Zhang W, Salcedo A, Rouse MN, Trick HN, Akhunov E, Dubcovsky J (2013b) Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. Science (80-) 783–786.

Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. Nat Rev Genet 3:429–441. doi: 10.1038/nrg817

Salmeron JM, Oldroyd GED, Rommens CMT, Scofield SR, Kim H, Lavelle DT, Dahlbeck D, Staskawicz BJ (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86:123–133. doi: 10.1016/S0092-8674(00)80083-5

Salmon A, Udall JA, Jeddeloh JA, Wendel J (2012) Targeted Capture of Homoeologous Coding and Noncoding Sequence in Polyploid Cotton. G3: Genes|Genomes|Genetics 2:921–930. doi: 10.1534/g3.112.003392

Samuels DC, Han L, Li J, Quanghu S, Clark TA, Shyr Y, Guo Y (2013) Finding the lost treasures in exome sequencing data. Trends Genet 29:593–599. doi: 10.1016/j.tig.2013.07.006

Schmutz J, Cannon SB, Schlueter J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. doi: 10.1038/nature08670

Schnable PS, Ware D, Fulton RS, et al (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science (80-) 326:1112–1115. doi: 10.1126/science.1178534

Schneeberger K (2014) Using next-generation sequencing to isolate mutant genes from forward genetic screens. Nat Rev Genet 15:662–676. doi: 10.1038/nrg3745

Schnurbusch T, Paillard S, Schori A, Messmer M, Schachermayr G, Winzeler M, Keller B (2004) Dissection of quantitative and durable leaf rust resistance in Swiss winter wheat reveals a major resistance QTL in the *Lr34* chromosomal region. Theor Appl Genet 108:477–484. doi: 10.1007/s00122-003-1444-4

Shatalina M, Wicker T, Buchmann JP, Oberhaensli S, Simková H, Doležel J, Keller B (2013) Genotype-specific SNP map based on whole chromosome 3B sequence information from wheat cultivars Arina and Forno. Plant Biotechnol J 11:23–32. doi: 10.1111/pbi.12003

Silva P, Calvo-Salazar V, Condón F, Quincke M, Pritsch C, Gutiérrez L, Castro A, Herrera-Foessel S, von Zitzewitz J, Germán S (2015) Effects and interactions of genes *Lr34*, Lr68 and *Sr2* on wheat leaf rust adult plant resistance in Uruguay. Euphytica 204:599–608. doi: 10.1007/s10681-014-1343-6

Singh A, Pandey MP, Singh AK, Knox RE, Ammar K, Clarke JM, Clarke FR, Singh RP, Pozniak CJ, DePauw RM, McCallum BD, Cuthbert RD, Randhawa HS, Fetch TG (2013a) Identification and mapping of leaf, stem and stripe rust resistance quantitative trait loci and their interactions in durum wheat. Mol Breed 31:405–418. doi: 10.1007/s11032-012-9798-4

Singh D, Mohler V, Park RF (2013b) Discovery, characterisation and mapping of wheat leaf rust resistance gene *Lr71*. Euphytica 190:131–136. doi: 10.1007/s10681-012-0786-x

Singh D, Simmonds J, Park RF, Bariana HS, Snape JW (2009) Inheritance and QTL mapping of leaf rust resistance in the European winter wheat cultivar "Beaver." Euphytica 169:253–261. doi: 10.1007/s10681-009-9959-7

Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Bhavani S, Njau P, Herrera-Foessel S, Singh PK, Singh S, Govindan V (2011) The Emergence of Ug99 Races of the Stem Rust Fungus is a Threat to World Wheat Production. Annu Rev Phytopathol 49:465–481. doi: 10.1146/annurev-phyto-072910-095423

Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Njau P, Wanyera R, Herrera-Foessel S a., Ward RW (2008) Will Stem Rust Destroy the World's Wheat Crop? Adv Agron 98:271–309. doi: 10.1016/S0065-2113(08)00205-8

Singh RP, Huerta-Espino J, Rajaram S (2000) Achieving Near-immunity to Leaf and Stripe Rusts in Wheat by Combining Slow Rusting Resistance Genes. Acta Phytopathol Entomol Hungarica 35:133–139.

Singh RP, Huerta-Espino J, William HM (2005) Genetics and breeding for durable resistance to leaf and stripe rusts in wheat. Turkish J Agric For 29:121–127.

Singh RP, Mujeeb-Kazi A, Huerta-Espino J (1998) *Lr46*: a gene conferring slow-rusting resistance to leaf rust in wheat. Phytopathology 88:890–894. doi: 10.1094/PHYTO.1998.88.9.890

Singh RP, Rajaram S (1991) Resistance to *Puccinia recondita* f. sp. *tritici* in 50 Mexican Bread Wheat Cultivars. Crop Sci 31:1472. doi: 10.2135/cropsci1991.0011183X003100060016x

Singh RP, Rajaram S (1992) Genetics of adult-plant resistance of leaf rust in "Frontana" and three CIMMYT wheats. Genome 35:24–31.

Smalle J, Vierstra RD (2004) The ubiquitin 26S proteasome proteolytic pathway. Annu Rev Plant Biol 55:555–590. doi: 10.1146/annurev.arplant.55.031903.141801

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46:431–437. doi: 10.1139/g03-027

Song W, Wang G, Chen L, Kim H, Pi L, Hoisten T, Gardner J, Wang B, Zhai W, Zhu L, Fauquet C, Ronaldt P, Disas R, Gene R (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science (80-) 270:1804–1806.

Sonnhammer EL, Durbin R (1996) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:1–10. doi: 10.1016/0378-1119(95)00714-8

Stein N, Herren G, Keller B (2001) A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. Plant Breed 356:354–356.

Swiderski MR, Innes RW (2001) The Arabidopsis *PBS1* resistance gene encodes a member of a novel protein kinase subfamily. Plant J 26:101–112.

Talajoor M, Jin Y, Wan A, Chen X, Bhavani S, Tabe L, Lagudah E, Huang L (2015) Specificity of a Rust Resistance Suppressor on 7DL in the Spring Wheat Cultivar Canthatch. Phytopathology 105:477–81. doi: 10.1094/PHYTO-09-14-0261-R

Teer JK, Mullikin JC (2010) Exome sequencing: The sweet spot before whole genomes. Hum Mol Genet 19:145–151. doi: 10.1093/hmg/ddq333

The International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science (80-) 345:12517881–125178811. doi: 10.1126/science.1251788

The TT, Baker EP (1975) Basic studies relating to the transference of genetic characters from *Triticum monococcum* L. to hexaploid wheat. Aust J Biol Sci 28:189–199.

Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. Proc Natl Acad Sci 108:20260–20264. doi: 10.1073/pnas.1116437108

Trick M, Adamski NM, Mugford SG, Jiang C-C, Febrer M, Uauy C (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biol 12:14. doi: 10.1186/1471-2229-12-14

Uauy C, Paraiso F, Colasuonno P, Tran RK, Tsai H, Berardi S, Comai L, Dubcovsky J (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biol 9:115. doi: 10.1186/1471-2229-9-115

van Poecke RMP, Maccaferri M, Tang J, Truong HT, Janssen A, van Orsouw NJ, Salvi S, Sanguineti MC, Tuberosa R, van der Vossen EAG (2013) Sequence-based SNP genotyping in durum wheat. Plant Biotechnol J 11:809–817. doi: 10.1111/pbi.12072

Vierstra RD (2009) The ubiquitin–26S proteasome system at the nexus of plant biology. Nat Rev Mol Cell Biol 10:385–397. doi: 10.1038/nrm2688

Vogel JP, Garvin DF, Mockler TC, et al (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768. doi: 10.1038/nature08747

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78. doi: 10.1093/jhered/93.1.77

Vrána J, Kubaláková M, Číhalíková J, Valárik M, Doležel J (2015) Preparation of sub-genomic fractions enriched for particular chromosomes in polyploid wheat. Biol Plant 59:445–455. doi: 10.1007/s10535-015-0522-1

Vrána J, Kubaláková M, Simková H, Číhalíková J, Lysák M a, Dolezel J (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics 156:2033–2041.

Wang HY, Liu DQ, Yang WX (2011) A wheat disease resistance gene analog of the NBS-LRR class : identification and analysis. 118:63–68.

Wang S, Wong D, Forrest K, et al (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796. doi: 10.1111/pbi.12183

Wanjugi H, Coleman-Derr D, Huo N, Kianian SF, Luo MC, Wu J, Anderson O, Gu YQ (2009) Rapid development of PCR-based genome-specific repetitive DNA junction markers in wheat. Genome 52:576–587. doi: 10.1139/G09-033

Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M (2015) Exome sequencing: current and future perspectives. G3: Genes|Genomes|Genetics 5:1543–1550. doi: 10.1534/g3.115.018564

Webb CA, Fellers JP (2006) Cereal rust fungi genomics and the pursuit of virulence and avirulence

factors. FEMS Microbiol Lett 264:1–7. doi: 10.1111/j.1574-6968.2006.00400.x

Wicker T, Matthews DE, Keller B (2002) TREP: a database for Triticeae repetitive elements. Trends Plant Sci 7:561–562. doi: 10.1016/S1360-1385(02)02372-5

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Simková H, Kubaláková M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Dolezel J, Keller B, Stein N (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 23:1706–1718. doi: 10.1105/tpc.111.086629

Wicker T, Taudien S, Houben A, Keller B, Graner A, Platzer M, Stein N (2009) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J 59:712–722. doi: 10.1111/j.1365-313X.2009.03911.x

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. Plant Cell 15:1186–1197. doi: 10.1105/tpc.011023

William HM, Hoisington D, Singh RP, González-de-León D (1997) Detection of quantitative trait loci associated with leaf rust resistance in bread wheat. Genome 40:253–260.

Williams ND, Miller JD, Klindworth DL (1992) Induced mutations of a genetic suppressor of resistance to wheat stem rust. Crop Sci 32:612–616.

Winfield MO, Wilkinson PA, Allen AM, Barker GLA, Coghill JA, Burridge A, Hall A, Brenchley RC, D'Amore R, Hall N, Bevan MW, Richmond T, Gerhardt DJ, Jeddeloh JA, Edwards KJ (2012) Targeted re-sequencing of the allohexaploid wheat exome. Plant Biotechnol J 10:733–742. doi: 10.1111/j.1467-7652.2012.00713.x

Winzeler M, Mesterhazy A, Park RF, Bartos P, Csosz M, Goyeau H, Ittu M, Jones E, Loschenberger, F. Manninger K, Pasquini M, Richter K, Rubiales D, Schachermayr, G. Strzembicka A, Trottet M, Unger O, Vida G, Walther U (2000) Resistance of European winter wheat germplasm to leaf rust. Agronomie 20:783–792.

Wu L, Xia X, Rosewarne GM, Zhu H, Li S, Zhang Z, He Z (2015) Stripe rust resistance gene *Yr18* and its suppressor gene in Chinese wheat landraces. Plant Breed. doi: 10.1111/pbr.12311

Xie W, Feng Q, Yu H, Huang X, Zhao Q, Xing Y, Yu S, Han B, Zhang Q (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci 107:10578–10583. doi: 10.1073/pnas.1005931107

Xu X, Pan S, Cheng S, et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195. doi: 10.1038/nature10158

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37:528–538. doi: 10.1046/j.1365-313X.2003.01977.x

Zadoks J, Chang T, Konzak C (1974) A decimal code for the growth stages of cereals. Weed Res 14:415–421.

Zhang L, Wang S, Li H, Deng Q, Zheng A, Li S, Li P, Li Z, Wang J (2010) Effects of missing marker and segregation distortion on QTL mapping in F2 populations. Theor Appl Genet 121:1071–1082. doi: 10.1007/s00122-010-1372-z

Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun

2:467. doi: 10.1038/ncomms1467

Appendix A

Virulence/Avirulence formulae of the *Puccinia triticina* pathotypes used to test parental lines in the field in Australia

Pathotype	Virulence/ Avirulence formula[a]	PBIC culture number
10-1,3,9,10,11, 12	*Lr1, Lr13, Lr14a, Lr16, Lr17b, Lr20, Lr26 / Lr3, Lr9, Lr15, Lr17a, Lr19, Lr21, Lr23, Lr24, (Lr27+ Lr31), Lr28, Lr35, Lr37, Lr42, Lr47*	592
76-1,3,5,7,9,10,12+*Lr37*	*Lr3, Lr13, Lr14a, Lr17a, Lr17b, Lr20, Lr26, Lr37 / Lr1, Lr9, L15, Lr19, Lr21, Lr23, Lr24, (Lr27+ Lr31), Lr28, Lr35, Lr42, Lr47*	621

Appendix B

Virulence/avirulence formulae of the seven *Puccinia triticina* pathotypes

Virulence/Avirulence formulae of the seven *Puccinia triticina* pathotypes used to test parental lines at seedling stage in the greenhouse in Zurich

Lr Isolate	temperature oC (night/day)	Lr genes																																				
		1	2a	2b	2c	3a	3bg	3ka	9	10	11	12	13	14a	14b	15	16	17	18	19	20	21	22a	22b	23	24	25	26	27+ 31	28	29	30	32	33	34	35	37	38	38	44	44
90035	16/20 n/d					v	i	v						v		i																								
		a							a											a						a				a	a									
91047	16/20 n/d		v	v	i	v	v	v		v	v			v	v		v	v	v			v			v		v		v			v		v	v						
		a							a							a				a	a					a		a			a	a									
93012	16/20 n/d		v	v	v			v		v	v			v	v	v	v		v						a	a	a	a	v		a	a	a		v						
		a				a	a		a									a		a	a	a			a									a							
95001	16/20 n/d		v	v	v	v	v	v		v	v			v	v	i	v	v	v			v			v					v			v	v	v	v					
		a							a											a	a					a	a	a		a		a									
95219	16/20 n/d			v	v	v	v	v		v	v			v	v		v	v	v		v	v			v					v	v		v	v	v	v					
		a	a						a							a				a						a	a	a		a											
96002	10/10 n/d		v	v	v	v	i	v		v	v	v	v	i	v		v	v				v	v	v	v				v			v	v	v	v	v	v		v	v	
		a							a							a			a	a	a					a	a	a		a	a							a			a
	25/25 n/d		v	v	v	v	v			v	v	v	v	v	v		v		v		v	v	v	v					v		v	v	i	v	v	v	v			i	v
		a						a	a							a		a		a					a	a	a	a		a								a	a		
96209	10/10 n/d		v	v	v	v	v	v		v	v	v	v	v	v	v	v	v				v	v	v	v	v			v	v		v	v	v	v	v	v			v	
		a							a										a	a						a	a			a								a	a		a
	25/25 n/d				v	v	v	v			v	v	v	v	v	v	v	v	v	v		v	v	v	v					v		v	v					v				v
		a	a	a					a	a										a					a	a	a	a		a			a	a	a	a		a	a	a	

a	avirulent response
v	virulent response
i	intermediate response
blank	not tested
in red	influence of temperature to response

Appendix C

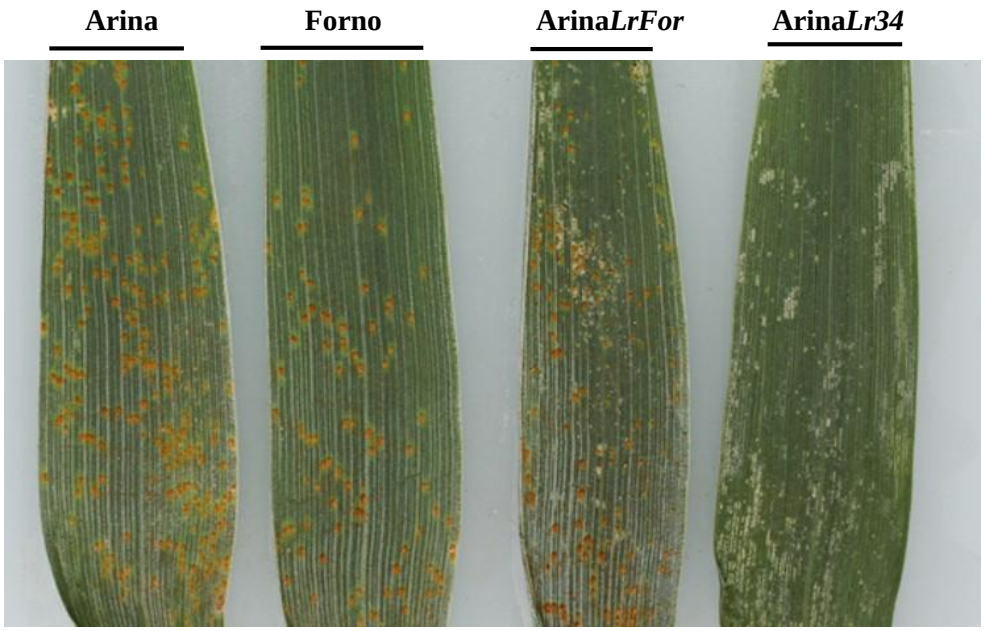
Sequence and repeat motifs of designed SSR markers along with the 1BS wheat Illumina sequence contigs

Primer name	Forward primer	Reverse primer	Repeat motif	1BS Illumina contig
swm201	CAGACGCCACCACCTTCCTC	TCCTCTGGGCCAGTGTCTCC	GAG(9times)	1BS_c-3469099
swm202	CTCAGACGCCACCACCTTCC	ATCGGGCCCTCTTCATCCTC	GAG(9times)	1BS_c-3469099
swm203	GCCTTCCCCATCAACCACAC	CCTCCGAAGAAGTGCCTGGA	GGA(6times)	1BS_c-3469099
swm204	GACGTCCCTCGCCTTCTTGA	TCCCCCTTCCACAAACATGG	AT(6times)	1BS_c-3484086
swm205	AGACGTCCCTCGCCTTCTTG	TCCCCCTTCCACAAACATGG	AT(6times)	1BS_c-3484087
swm206	CAGGTTGCACGTGGAGATGG	CGACGACGGTTTCTCGTCCT	AAT(6times)	1BS_c-3484088
swm207	ACGAGCTCGGAGCCATGTTC	GCGGCGTAAGGAGGAAGAAG	AAT(6times)	1BS_c-3484089
swm208	ACTGGGCAAACCAGCGATGA	AACCAGTGCATCACACCACATC	AT(6times)	1BS_c-3484090
swm209	TGGGCAAACCAGCGATGAAC	AACCAGTGCATCACACCACATC	AT(6times)	1BS_c-3484091
swm210	TGGCCATTGATGTTAATCTGATGC	GAAACCAGTTAGGGCGTGAAAG	TA(6times)	1BS_c-3440731
swm211	GGGCAGAGCACCAATGGAAG	GCAGGCACACCATCATCACC	AG(6times)	1BS_c-3440732
swm212	TGCAGCAGACGAGGAGCTTG	GGGCGCAGATGAACAAGATG	CA(8times)	1BS_c-3416697
swm213	CGGTACCGGCGTCAGAAAGA	CCTACCGCTTCATCGCCAAC	GT(11times)	1BS_c-3480048
swm214	TACCGGCGTCAGAAAGATCG	CCTACCGCTTCATCGCCAAC	GT(11times)	1BS_c-3480048
swm215	GATGTGCGGCAAAACACACC	GCCTCCGAGACGTGACAATG	GTT(7times)	1BS_c-3480048
swm216	CAGCCCCGAATAATCGAGCA	GCTTGGGCTTTTCGGCTACC	AG(15times)	1BS_c-3412394
swm217	CCAGCAGCAGAATGCCTTGA	ATGTGCAGCAAAGCGCACTC	CAA(15times)	1BS_c-3476654
swm218	TCCAGCAGCAGAATGCCTTG	ATGATGCGGGTCCATCAAGC	CAA(15times)	1BS_c-3476654
swm219	CCCATGCTCTTCACGTGCTC	GCCAGAGAACGCCAAACCTC	GA(9times)	1BS_c-3475803
swm220	AAGGCAACAAATGCCCATGC	ATGTGCTGCCTCGGTAGTCG	GA(9times)	1BS_c-3475803
swm221	CGCTGCTATTCGAGTGAGCG	TGCTCATGTCATCCCACGTG	GA (7times)	1BS_c-3436438
swm222	GTCCTCGTCGCTGCTATTCG	CCTTGCTCATGTCATCCCAC	GA (7times)	1BS_c-3436438
swm223	GCTTCCTCCAACGCCATCTC	TCATCCCACGTGCATTCTCA	GA(7times)	1BS_c-3436438
swm224	TAGTCGAGCACTGCAGGTAG	TCCACGAGAAGAGAGTGGAG	CA (7 times)	1BS_c-3477533
swm225	AGCACTGCAGGTAGGTATCC	TCCACGAGAAGAGAGTGGAG	CA (7 times)	1BS_c-3477533
swm226	TCACGTTCCAGTAGGCTAGC	TCCACGAGAAGAGAGTGGAG	CA (7 times)	1BS_c-3477533
swm227	CGATGCCTAATCGGCTGGAG	TGCCGCGTGTGTTTGAAATC	TC(6times)	1BS_c-3432548
swm228	AGTGGAATCCGGATCAGTGG	TGCATGGTCTGTCCAGCATG	CT (9 times)	1BS_c-3471491
swm229	ACTGCTAGTAGTCGGCTCGG	ACTCTGCATGGTCTGTCCAG	CT (9 times)	1BS_c-3471491
swm230	GCCCTGAGAAGGTCCATTCG	CCATGTGCCAGAGCCATGAG	AAC(7times)	1BS_c-3418853
swm231	GAGCTGGTGCGATGAACTAC	GTCTAACCGCCACAGAGTAG	CT (7 times)	1BS_c-3482242
swm232	CTCAGGGGCGGATGATGAAT	GGATGGAGGTCACCGTGGAG	CT(7times)	1BS_c-3482242
swm233	CCGAGAGACTCAATCCATGC	GGGAGCGTCACAAGAAGTAG	AT (8 times)	1BS_c-3482926
swm234	ATGGTTGGACGAACGACAGG	TTGGATCAAGCGGCCACTGG	AT (7 times)	1BS_c-3456182
swm235	TCAGATGCCTCCGAGACGTG	CTCCGCATTTGTGCATCCTG	ACC(6times)	1BS_c-3464043
swm236	CCTCGATGAAGCGTGAGGAG	CGCAATACCGTAACCGCATG	AG (6 times)	1BS_c-3420584
swm237	AACGCCGCCTGGTAAGCTG	CTGCCTCCACCCATTTGGTC	CTC(7times)	1BS_c-3436589
swm238	TCCTCACCTTCCCTCCGTTG	GACGTGACAGAGTCCCTAAC	TC (11 times)	1BS_c-3415016
swm239	AGACCCCTCTTGGCCACGTC	TGGACGTGACAGAGTCCCTAA	TC(11times)	1BS_c-3415016
swm240	AGCCAGGCAGTTGCTGCATG	TGGCCACAAGCAGTCTCATG	TC (8 times)	1BS_c-3484126
swm241	CGTACGGCAGTGGCGAAGTC	TCCAAACAGTGGCCACAAGC	TC(8times)	1BS_c-3484126
swm242	AGTGGGCTGCGGTAGGTTAG	TCCAAACAGTGGCCACAAGC	TC(8times)	1BS_c-3484126
swm243	CCAAGTTCGATCCCTCACG	TTATGACCCGGCAGATGCAA	TGA(6times)	1BS_c-3480724
swm244	ACAAGCTTCAGGTCGAGGAG	ATGCCTCGTCTTGCTGCTGC	TG (16 times)	1BS_c-3431246
swm245	CCTCTCCGATGATGCGTGAGT	CAAAGCACCTCTTCCCATTGC	CT(14times)	1BS_c-3437094
swm246	GGCCTGATCCCCAAGATGTG	CCTCGATGCTGGCCCTACAC	TG(16times)	1BS_c-3437094
swm247	GCCTGATCCCCAAGATGTGC	CCTCGATGCTGGCCCTACAC	TG(16times)	1BS_c-3437094
swm248	CCATCTTCATCATCCCGGTG	CGTGATCCCATTGTCACCAC	TC (7 times)	1BS_c-3456100
swm249	CATCTTCATCATCCCGGTGC	CATCAAGTGGCACGTGATCC	TC (7 times)	1BS_c-3456100
swm250	AGGGTTCAGGCCTCGTCCTC	CACCGCGATGCAAGTTTGAG	CA(13times)	1BS_c-3481704
swm251	CAATCCCACCGGTCGGAATG	GCGGTACTACCATGATGACC	GA (7 times)	1BS_c-3455507
swm252	ACCGGTCGGAATGTGAGGAG	GGAGATGGGTTGCCTTGTGC	GA(7times)	1BS_c-3455507
swm253	TCGACGAGTCGACGGTGTTC	GCCTCTCCCCCTGAAATTCG	GA(6times)	1BS_c-3417129

swm254	GAATCGAACGGCTCTGATGC	TCATGCACTACCCAGCCCAC	GA (7 times)	1BS_c-3446367
swm255	GACATGGTGTTGCGGTCAGG	GGCGAGTGCGAGAGGCTAAC	GAGG(6times)	1BS_c-3446367
swm256	TGTCCTCCATTGCCTCGTCG	TCGACTCCGAAGACATCGAC	AT (5 times)	1BS_c-3477508
swm257	AGACGTATCAGGTCGTCGTG	CTAGGTCAGTCCATCCAAGC	CA (10 times)	1BS_c-3477105
swm258	CTCTCCGCATCTGTGCATCC	TCAGATGCCTCCGAGACGTG	GTT(8times)	1BS_c-3423367
swm259	GGCAGCCGGAGCTATGTGTG	CGTCATCGAGGAATGGCTGA	AG(8times)	1BS_c-3436117
swm260	ACTTCGCAGCCTGCCCACTA	GGTAGCCCTCTTGCGTGTGC	AG(17times)	1BS_c-3436117
swm261	ATGTGCTGCATGCGTGTGTG	CACGCAACATCAGTGCGAGG	TG (7 times)	1BS_c-3483722
swm262	TTGGCACGTCTTGTCGAAGC	TCTTGCGGGCAGAGAAGGAG	TC(25times)	1BS_c-3478381
swm263	TGCATCTCACCAACCAGTGC	AGCATCTGGTACACTGGCTG	CA (10), GA (14)	1BS_c-3468881
swm264	CCTCCAAGATGGGGACAAGG	GGCCAAGCCCTTCTCAATCC	GA(14times)	1BS_c-3468881
swm265	ACTGCTCGAGCTGCTTGGTC	TAGCCTCACGTACGCAAGTG	AT (8 times)	1BS_c-3470291
swm266	GCGCTACTGCTCGGGAACAC	CGCCCAGAAACGAGAGGATG	GT(9times)	1BS_c-3444735
swm267	CAGGGTCTCCACATGATCAC	GCCTCCATGGACACTGACAC	CT (6 times)	1BS_c-3414083
swm268	GTTCCGTGCAATGCATGAGG	TGCACACCATCCCGTGTACC	CA(8times)	1BS_c-3482914
swm269	TGAATCGCCACGAGTACGAC	GTTAACCAGGATGCTGGCTC	CT (7 times)	1BS_c-3420667
swm270	GCCCAAGCGACCTGCTACAC	GGAGAGAGCCGCAGATGGAG	CT(6times)	1BS_c-3453805
swm271	GTCCATTCGGCGCTAGATCG	CTGGCTCCGGCACCTTATCA	CT(7times)	1BS_c-3420667
swm272	GCCCAAACGACAACATGCAC	TGCTTCAACGTCGCAGTGTG	GA(12times)	1BS_c-3462795
swm273	TTGTCCCTGGTGCACTATGC	TCACCAGGTCTTGCCATCTC	CT (9 times)	1BS_c-3450437
swm274	TTGCTCATGTCGTCCCACCT	GGCATCACGGGCTTCATAGG	CT(12times)	1BS_c-3467166
swm275	GTGGGCGCAGTAGCAAAGGT	GTCGTGATGCCGTGGTCAAG	CT(12times)	1BS_c-3429013
swm276	GGCCAGTCATGTACTTGGTTTCA	TCTACGCCTCATCCCGCTTC	GAA(24times)	1BS_c-3461491
swm277	TAGGAGAGCTGCCCTGATCG	ATGCACCACCACCAGCTTGA	CT(8times)	1BS_c-3450437
swm278	AGTAGGAGCGGGGGTGGTTG	TTGACTTCGGAGTGGCACGA	TGT(17times)	1BS_c-3450437
swm279	CACATGTACGTGCTGCTTCC	CCAACCTCCACGACGTGATC	AG (19 times)	1BS_c-3481252
swm280	TGCATCCAACCCTAAAAGCACA	TATCCCCTCCGACCGTATTCAC	AT(10times)	1BS_c-3479571
swm281	TCCGAAGCGTATGCCATGCTA	TCAACTGACGATGGCCTCTCC	TC(7times)	1BS_c-3467719
swm282	TGCTTCCAAGGATTGCCTCA	GACGACCCCGCAGTGGAC	AG(19times)	1BS_c-3481252
swm283	CAGCCACTCGGGTTTTGACC	CAAGGGTCCAAATCCAACAAA	AG(7times)	1BS_c-3482538
swm284	TTTCAGCCGCCACAAGTTCC	CCCAACAAACACACCGCAAA	CT(20times)	1BS_c-3443776
swm285	GACGCTGGGGCAAATAGCAC	GGCGAAAAGCCACTGTCTCG	GA(7times)	1BS_c-3437609
swm286	CACGGACGTGCACCAGTCTC	GCGCAAGTGCTCAATGCTGT	AC(6times)	1BS_c-3470582
swm287	TCAGATGCCTCCGAGACGTG	CTTGCGGAGCCAATCATCTG	AAC(12times)	1BS_c-3434216
swm288	GCCCATAGCGGTGTGAGAGC	TGTGCAAACCAACACGAGCAC	TG(10times)	1BS_c-3480776
swm289	TGCGGGTGTCATGTCTTTCA	TCCCTCTCTCCTGCAAACCA	GA(21times)	1BS_c-3462770
swm290	TTGGAAGCTGAACGGGCATT	TTACGCACAGCCGACACTGG	GGA(6times)	1BS_c-3484800
swm291	GGGCCTTCATCAGTGTGTTGC	TGATGGCTGGTGTTTGGTTTG	CAG(6times)	1BS_c-3463234
swm292	CCAAGGGCATGCCTGAAGTC	TTGGCTGCCATGTCACACCT	TC(7times)	1BS_c-3457640
swm293	CGCTCTCCTCTCCGTTCACC	GTCCACACAGCCAAGCCAAG	CT(22times)	1BS_c-3457640
swm294	GGAGGCCCAACATGAGTCTA	AGTCCACCCTTGGGGCTGAG	AG(10times)	1BS_c-3465657
swm295	TGGGCGTTCGAGTCCGTATT	CGAACAGGCTGAACGGGAAG	CTT(7times)	1BS_c-3472946
swm296	TGACGTGCACTTGCGTTCAG	TCTGGCTTCCGCCAGAGAAG	TA(7times)	1BS_c-3415373
swm297	AGGGCAACATCCGACTACGG	GCTTGCTGCAACCTCACCAC	GAA(7times)	1BS_c-3445460
swm298	GCATCAGTTCAACCGCTTCG	GGCCTCTAGGCCTCCCAACC	TA(8times)	1BS_c-3427031

Appendix D

Disease severity of parental lines at adult plant stage in the field in Australia



Appendix E

Primer sequences for developed KASP markers for genetic mapping

KASP_name	Primer_Allele Forno	Primer_Allele Arina	Primer Common	Allele Forno	Allele Arina
B2g40930	CTGTGTAGGCTAGCTCGAAATACTCTA	GTAGGCTAGCTCGAAATACTCTG	GCTCTGCGGCCGCAACTGTATT	A	G
B2g40820	CTAAACAAGTGTTGTCTTTCTAAACAGGC	CAAACAAGTGTTGTCTTTCTAAACAGGT	CGTGCTTTGTATAGACAAATGAAAGCCAT	C	T
B2g40480	CTGCCTCACCAGCATAATCAATTTCG	CGCCTCACCAGCATAATCAATTTCA	AAAACTGGGTCTAGATCAGGTCACTAAAA	G	A
B2g40060	CTCAGCTCCTGGAGCAGGGGT	AGCTCCTGGAGCAGGGGC	CCTGAGCCTCCTCAAGCACCAA	A	G
B2g39790	CTGGCAGCAATGTAATTATGGAGTAGGA	GCAGCAATGTAATTATGGAGTAGGC	AACACTACTACAAGGTTTCCATCGGTTT	A	C
B2g39710	CTGCAGCATGTAGTGATATCTGACG	CGCAGCATGTAGTGATATCTGACA	AAATAAGTGAATGTGGATCACATTCACCTT	C	T
B2g39700	CTATTCATGCTAGTTCCATGAAGTTCTG	CTATTCATGCTAGTTCCATGAAGTTCTA	CTCAGTGGATCAAGTATGTCAGGGAT	C	T
B2g39420	CTAGTTATATTTTACTTTGTACAAAACCTTATTCC	AGTTATATTTTACTTTGTACAAAACCTTATTCT	TCATAGTTACGAAGCCTTGCTTTTCTGAA	G	A
B2g39390	CTCCTTGGAATGCTGCTAAAGAGGG	ACCTTGGAATGCTGCTAAAGAGGA	CTAAAAAGGCCTTAGGCCCGGCTT	G	A
B2g39360	CTAACTGAATAAAATGCCATAATTACCTAAATAT	AACTGAATAAAATGCCATAATTACCTAAATAG	GTAACATGATATCTGCACGCACTTACAA	A	C
B2g39290	CTACTATTGGAGCTGGAAGAACATCG	GACTATTGGAGCTGGAAGAACATCA	CCATTGCACGCGAACTGAAAGAGTA	C	T
B2g39280	CTGAGTATCCCGTCTAGCCCTTGT	GAGTATCCCGTCTAGCCCTTGA	TTTACCAATGCTAACTAGGTAAAGATCGAT	T	A
B2g39180	CTGTGCATCATACAGTATGATTCACCC	GGTGCATCATACAGTATGATTCACCT	GCTTAAGGTACCTGTGAACATATTCAGAA	C	T
B2g38910	CTAAATAACTACATGATTGAGAATCTCACAG	AATAAATAACTACATGATTGAGAATCTCACAT	GACAGCACGTTGAATACCTCTTTGTATAA	G	T
B2g38850	CTCAATTTTTTGTGATTTTTTTCTGTAGTGTAGA	CAATTTTTTGTGATTTTTTTCTGTAGTGTAGG	GCCTCCAACGGTCCACACGAAA	A	G
B2g38480	CTTTTATCACAATGTAAGCAAATTGGGTG	CTTTTATCACAATGTAAGCAAATTGGGTT	AAAGAAAGAAACACATCCTCCGTTGACAT	G	T
B2g38440	CTACCAACAGTACACCAAATAACAGCAAA	CCAACAGTACACCAAATAACAGCAAG	CCTCCATTTAGTTCTGCCCATGGTT	A	G
B2g37790	CTAGATTTCTCATGATTGTCCCTTACAT	AGATTTCTCATGATTGTCCCTTACAC	TCTCTGGGTGCGGCCAGAAGTT	A	G
B2g37770	CTGAAAAACATATTCCAGAACATGTACTGAC	GAAAAACATATTCCAGAACATGTACTGAA	CTTTCTTGTGGAGAGTGCTGACTGAA	G	T
B2g37750	CTATGAGGAAAAATTATGCTATGCCCTGT	GAGGAAAAATTATGCTATGCCCTGC	CCCTTCTTTCTTCCATTGACAGTAAGTT	A	G
B2g37190	CTCATCAACAAAATAAGACAGTTCCTCAA	CATCAACAAAATAAGACAGTTCCTCAG	TGATGACACCACATGCATATTAGACCAAA	T	C
B2g37070	CTGACATGGAATGACCGACGGTAAC	GACATGGAATGACCGACGGTAAT	CCAGCTTATAGTGCCAGCCAATAGTT	C	T
B2g37060	CTATCAAATAACATGAAAGAACAAAAAACACATG	CATCAAATAACATGAAAGAACAAAAAACACATA	GACTTGTCATTGAAATACCTAGATATGGAT	C	T
B2g37000	CTATGCCAAGGATGGCAGTTGCAC	ATGCCAAGGATGGCAGTTGCAT	GAATGCGCCAGGTACCATGGAGAA	C	T
B2g36930	CTATATTTATGCTGTAGTGATTAACCATGTGT	TATGCTGTAGTGATTAACCATGTGC	ATACAAGGTACAACTTGTTGGCAGTGTT	T	C
B2g36920	CTCATCATCCTCCACTTTCATGGACT	CATCATCCTCCACTTTCATGGACA	AATAGTAGGAGGACAAAGACGATCTAGAT	A	T
B2g36910	CTTGACAATTAAAGTTAGCATTCTCGC	ATACTTGACAATTAAAGTTAGCATTCTCGT	CATGGCAACTCTGAACTTTTTTATGGCAA	G	A
B2g36820	CTGAATAGCTATTGATGAAGAGTATGTGC	CAGAATAGCTATTGATGAAGAGTATGTGA	TGCCTGGTCCGGTTTTCCTTTGAAA	G	T
B2g36760	CTATACAAAGCTCTCAAGGATTAAAGGAC	CATACAAAGCTCTCAAGGATTAAAGGAT	AACATTACAGATCCTGAAGTTGGGGTAA	G	A
B2g36730	CTGGGCTCCATTTCCACGATAAT	CTGGGCTCCATTTCCACGATAAC	CGAATTAAAAGGGGATCATGGAAAGTGAT	T	C
B2g36490	CTACCAACATGGTTTAGACATACACCG	CACCAACATGGTTTAGACATACACCA	TGTTTCCATGATAAGCAGCTAGGKTCT	G	A
B2g36400	CTCATAAGAAAGGCTCCATCTGCTGTA	ATAAGAAAGGCTCCATCTGCTGTG	GATAAATCATGTACAAAGGGAACTAAGCAA	A	G
B2g36380	CTAGATGCCTACTGGCTTAAATAAATTATA	AGATGCCTACTGGCTTAAATAAATTATG	AGCAGATGATCTCGCAGAATCAAGAATT	A	G
B2g36330	CTTCTTAATGAAATCTTGGTCATAAACGTAG	CATTCTTAATGAAATCTTGGTCATAAACGTAT	CGTGCTTCAATGCAGCAACTAGATAATTA	G	T
B2g36160	CTTTTGTTGGGGACACTTGATATTTCATTTT	GTTGGGGACACTTGATATTTCATTTG	ATCAACCACATCTTCACCAACACTCTTT	A	C
B2g36150	CTGATCTGCAGTCTCTGTACAATCAG	ATGATCTGCAGTCTCTGTACAATCAA	CACAATTAAGCACTTCTTGCCACTAGAAA	C	T
B2g35930	CTGCAAATTCGGTGCGCCACTAGT	CAAATTCGGTGCGCCACTAGC	CTAGGAAAAATTTCCAATGCACAAGCCA	T	C
B2g35800	CTTATCCAAAACCTACAGGCAGCTG	CTTATCCAAAACCTACAGGCAGCTT	GATAAGCACGCACCAAACCTGAGAA	C	A
B2g35750	CTCATTTCAGAATTCATGGAAGGCATTCA	CATTTCAGAATTCATGGAAGGCATTCT	CTCAGTTAGTACTTTAGGGAAGTGGTAA	A	T
B2g35740	CTGCAGGATGTCATATCATGAACTTATGTAA	CAGGATGTCATATCATGAACTTATGTAG	AAATGCTAATAGCCATACGTAATCGACAAA	A	G
B2g35720	CTCAAAGCAAACCTTATACGCTGAAGAT	CAAAGCAAACCTTATACGCTGAAGAC	CGTGCTTCACAGGAACATCAGTGTA	A	G
B2g35690	CTGAAGTAATCAAGAGCTATCCAGC	CTGAAGTAATCAAGAGCTATCCAGT	GCACACTCAGCTATGAGATCTTTGATTT	C	T
B2g35620	CTGCCATGTCTTGCAAGAGGCGG	GCCATGTCTTGCAAGAGGCGA	TTTCCGTTAGATGAAACTCGGCAAACAAA	C	T
B2g35590	CTATTGATACACCAGGCCTTCGGC	GATACACCAGGCCTTCGGG	GCGAGAATTTTCCTGTTGGCTCCTT	G	C
B2g35150	CTACATATGGACTACAATAATACTAATACCG	GACATATGGACTACAATAATACTAATACCA	GACCAACACTGCTGTTCTTGGTTCTA	G	A
B2g35070	CTATTGATCTTTCCTTTTGTCCAACACTG	GATCTTTCCTTTTGTCCAACACTT	CCAAAGTGATGACCTTTGCTATTAGCAA	G	T
B2g35020	CTGTAAACCCTTTTACTGCTAATAAAGTGAAT	GTAAACCCTTTTACTGCTAATAAAGTGAAA	ACATCTGATATCTGTTTTCTCTGCCGTTA	T	A
B2g34970	CTGATGCCTGGTTTTGATCCTGC	CTGATGCCTGGTTTTGATCCTGA	GGATTGATAAGGACTCTCTGAGCCAA	G	T
B2g34880	CTGTGGTTGTGGAAGTTGTATTGGACT	GTGGTTGTGGAAGTTGTATTGGACA	CACCACTTTCAGATTGAGTGGGCTA	T	A
B2g34840	CTAGTTTTCCGAGTCACTAGATCCAAAA	GTTTTCCGAGTCACTAGATCCAAAG	CGTTTGCTGGCTATTGGTTTTTTATTTCTT	A	G
B2g34820	CTCAATATCATTACTGTCATCGTTAGCCA	AATATCATTACTGTCATCGTTAGCCC	ATAATATATCAGCAACAGCTCTGTCGTCT	T	G
B2g34710	CTACCACCAGACCAAACAAAGATCG	GACCACCAGACCAAACAAAGATCA	TGCCTTGGTTGGAGTCTCTTTGGTT	C	T
B2g34660	CTCAAGTGTCCAGCCATTACATGCC	ATCAAGTGTCCAGCCATTACATGCT	CAGAGACGTGCCTTGGCTGTCTA	C	T
B2g34610	CTATACATTTCCAGGATTTGGGTATCCAA	ACATTTCCAGGATTTGGGTATCCAG	CGTGCTCTTGAGAAAACGAGGGTTT	A	G
B2g34550	CTGTGTTTAGGCGGGCCTTAAAGG	GTGTTTAGGCGGGCCTTAAAGC	CATTCCTCAGGCCTTGCTAGTCTTT	G	C
B2g34490	CTGGTATCATGATATATGCTTACCACATCA	GTATCATGATATATGCTTACCACATCG	AGACTTTCAGGATGGACGGACC	A	G
B2g34450	CTGTTTCTATAACCATTACTGTTAAAACCCTA	CTATAACCATTACTGTTAAAACCCTG	TTCGGCTGAAAGGTTTGTGAGTTAACAAA	T	C
B2g34240	CTATAGAGAATGGAAGTCACGGTGC	ACATAGAGAATGGAAGTCACGGTGT	CCTTCATTTTACACGAGGCTAACATTGTA	G	A
B2g33570	CTGCATTGAGATTTCAATTGCAGGCTA	GCATTGAGATTTCAATTGCAGGCTG	CCATGGTGCTTGGGCTCGTCTT	A	G
B2g33060	CTGCAATTTGGTCCAAGTAAAATCCCC	GGCAATTTGGTCCAAGTAAAATCCCT	GATTGTCTCTTGCTCTTAATGATGCTCTT	G	A
B2g32910	CTCACTAGTACTACGTTACTTTGCACTATA	ACTAGTACTACGTTACTTTGCACTATG	GGGTCATCTATTTTGGAACGAAAGGAATA	A	G
B2g32520	CTCACTGGAGGGCATGCAGCAC	ACACTGGAGGGCATGCAGCAT	AGGAGCTCAAGCCAACGATTTGCAA	G	A
B2g32310	CTCACAAAATACAGGGAGCTGTACCA	ACAAAATACAGGGAGCTGTACCC	CGATATGCCGCTCCATAAAAACACTAAAA	T	G
B2g32300	CTCACTTCTGTATGAAATTCATTAGTAGCTAT	CACTTCTGTATGAAATTCATTAGTAGCTAA	TTGTTGGATATGTAACAATCTCTCAAGTTA	T	A
B2g31720	CTCTCTCCGAGGAGATTGAATGC	CTCTCTCCGAGGAGATTGAATGT	CCGAGGCCTAATTAACATTAGTAGACTTA	C	T
B2g31570	CTTCTACACGAGTGCTGAAATGCC	GTTCTACACGAGTGCTGAAATGCT	ATGCTAATGGACCGAGGAGGTCTTT	G	A
B2g31490	CTGAATGTAGTTTCCACTTCTATTATTTTGCTA	AATGTAGTTTCCACTTCTATTATTTTGCTG	CCTTGTATTTTGGGTCATTCTGATCATGAT	T	C
B2g31110	CTGCAACATCAAATTGAGTGAACCGTGT	CAACATCAAATTGAGTGAACCGTGC	GCTAACCTTAGATCACCCTTTTGGCAT	A	G
B2g31100	CTATACACCTTATATCAAGGAATGGAGGA	ACACCTTATATCAAGGAATGGAGGG	GAAGTGCATCGTAGGTCCTTGAACTA	A	G
B2g30170	CTGAGTTTCCCGTTTCCTTCGCATT	GAGTTTCCCGTTTCCTTCGCATC	GTCCCCCTACGGAAGATAGAACATA	A	G
B2g28450	CTAAGCATTGGTATAACTGCACCACTAT	AGCATTGGTATAACTGCACCACTAC	AATGGTAGCCAGAGCTTTTAAGCTGTT	A	G
B2g27400	CTACTACCTCCGTCCGGGATTAT	ACTACCTCCGTCCGGGATTAC	ATGCAACAACTCAATGCAGGGCTTTAAAA	A	G
B2g27350	CTGACAGTTGATACATGGTTAATATTTCCAG	GACAGTTGATACATGGTTAATATTTCCAC	AATGCCCAGTAAACCGAGTCAGGAT	C	G
B2g27110	CTAGGCAGCACCAGAAGTAGCG	CTAGGCAGCACCAGAAGTAGCA	GGCCTGGCCCATTTGTGGACAA	G	A
B2g26990	CTGCAATGGAGTACAGAACAACTTGC	GCAATGGAGTACAGAACAACTTGG	GTTCAACTTTGGTGCTGTAACCTAATGAA	G	C
B2g26890	CTCCATCGGCAATTAATATGGATCGAGA	CATCGGCAATTAATATGGATCGAGG	CCGCCCACTAATCTAACTGTTAGAATATA	T	C
B2g26840	CTCTTGACTGACAGCCTATTGTACC	CTCTTGACTGACAGCCTATTGTACT	AGGAAACTAAGTGAAACGGAGATCACTTT	C	T
B2g26790	CTGGCAAGCAAACGACACCGAATTG	AGGCAAGCAAACGACACCGAATTA	GTCCGTGCTTCATATTCTTACACACTTT	G	A
B2g26780	CTCCAAGCTCGGAGGATGCAGTTAT	CCAAGCTCGGAGGATGCAGTTAA	GATCGATCGAACGGTCTAAGTAGCAA	T	A
B2g26710	CTCAAAGGTTTAATGTCCAATGATATCTTGTT	AAAGGTTAATGTCCAATGATATCTTGTC	CCTTAGGCATCTGATTGTGCCAATTATAT	T	C
B2g26650	CTGATTCTGCTCTGATGCATACTAGC	GATTCTGCTCTGATGCATACTAGG	CAGAATCGTTGGGTTGGCAACTTTATAAT	C	G
B2g26600	CTGAGGATGGAAATGGTCTCCACG	AGAGGATGGAAATGGTCTCCACA	CCACTTTTCTGTACCACACAAGTAGTTTT	G	A
B2g26300	CTGGTCACGACACAATTCGTAACCTTAT	GTCACGACACAATTCGTAACCTTAC	CCCTTGTCTATCACTACATGCTAACATTT	T	C
B2g26260	CTGGCTTGTAACAACCAATCTATGGTATA	GCTTGTAACAACCAATCTATGGTATG	TATCCGTTTTCCTGAGGAAAAACGACAA	T	C
B2g26200	CTAACATGAACTGGTGCAGTATGCTG	CAACATGAACTGGTGCAGTATGCTA	CCTGAACTGCGAAGTCAGCTAAAGAT	C	T
B2g26180	CTATCATTGTAGCTCCATAGACC	CGCTATCATTGTAGCTCCATAGACA	TAAAGTCAGCGATGACCACGGGAA	C	A
B2g26170	CTCAGCTCCAGCTGTGTTTCTCTG	ACAGCTCCAGCTGTGTTTCTCTA	CAGACACCGTCAAGAAGACTGTAGAA	G	A
B2g26100	CTGGATGTCCATATTTCCCGAGAAGAAA	GATGTCCATATTTCCCGAGAAGAAG	CCTGTGGTCTTGCTGTTCTTGACAT	A	G
B2g26090	CTACGGAATCTTGGTTCGGCTCC	CACGGAATCTTGGTTCGGCTCT	ACTCGCAGCCGCGTCTGTTCTT	G	A
B3g19870	CTAAAGCATAGGTGGGTTTTCCAACAC	CAAAGCATAGGTGGGTTTTCCAACAT	AAAGATGGTATGCTCATATAAACCGCAAAA	G	A
B2g38640	CTCATTTTTTCAGGCGGTAGCCCG	CCATTTTTTCAGGCGGTAGCCCA	AGAAAAGCACAGGAAACAATGGGAGAAT	GG	AT
B2g36970	CTAAACAACTCATTGTGATAGCCATCGA	AACAACTCATTGTGATAGCCATCGG	CGACATTTGTTTAGTTCCCATTAATTAGTT	T	C
B2g36940	CTGATGGCGACAAGGAAACTGCAGT	ATGGCGACAAGGAAACTGCAGC	CGACAGAAGAAGAACCATTCAGTCCAT	T	C

Appendix F

List of wheat varieties used for marker diversity analysis

Wheat line	Country of origin	Wheat line	Country of origin
Amigo	USA	Mulan	GER
BWS26SUIline2	MEX	Nara	CH
Cambrena	CHN	Normandie	USA
Camedo	CH	Pavon76	MEX
Campala	CH	Rainer	AUT
Caphorn	UK	Renan	FRA
Chancellor	USA	Rubli	CH
Chineselandr34	CHN	Runal	CH
ChineseSpring	CHN	Rustic	FRA
Claro	CH	Scaletta	CH
Combin	CH	Segor	CH
Disponent	GER	Sertori	CH
Fiorina	CH	Siala	CH
Forel	CH	Suretta	CH
Galaxie	FRA	Tapidor	FRA
Glenlea	CAN	Thatcher	MEX
Impression	GER	Titlis	CH
Kanzler	GER	TP1142XStarke	SWE
Levis	CH	Transec	USA
Ludwig	AUT	Winnetou	GER
Manhatten	GER	Zinal	CH
MarisHuntsman	UK		

USA: United States of America, CAN: Canada, MEX: Mexico, CHN: China, UK: United Kingdom, CH; Switzerland, GER: Germany, FRA: France, AUT; Austria, SWE: Sweden

Appendix G

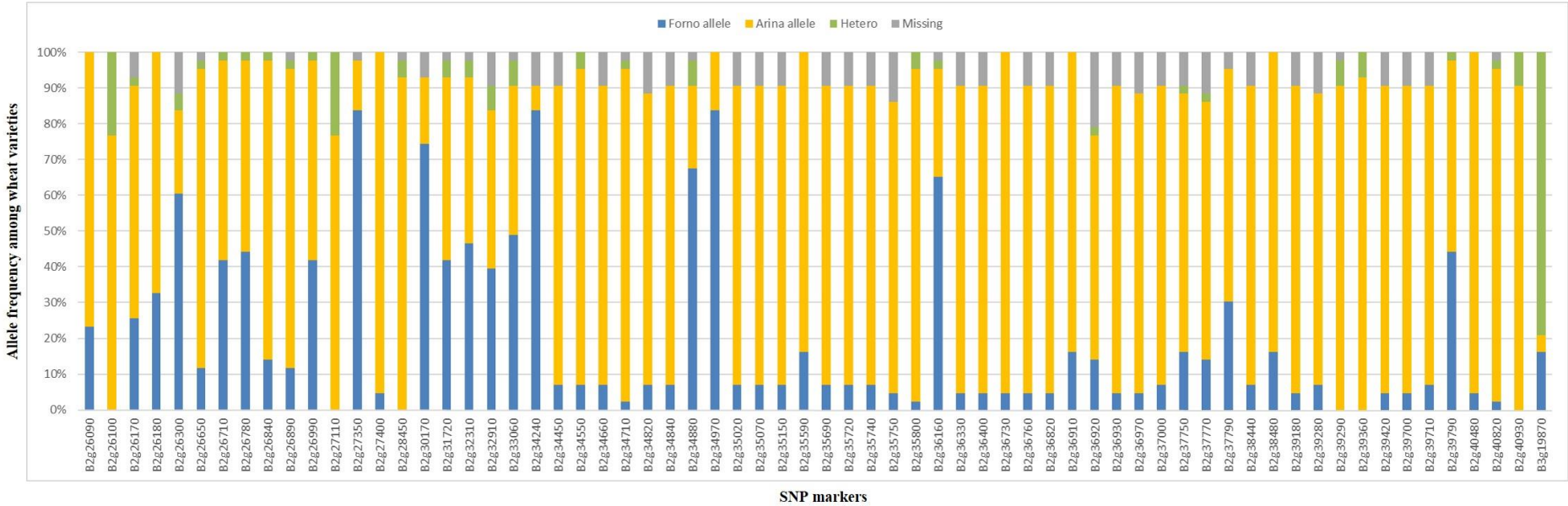
Analysis of the 80 SNP markers in 43 lines of wheat germplasm

Marker	Major Allele Frequency	Heterozygosity Frequency	PIC value	Forno allele	Arina allele	Heterozygous allele	Missing	total
B2g26090	0.77	0.00	0.36	10.00	33.00	0.00	0.00	43.00
B2g26100	0.88	0.23	0.36	0.00	33.00	10.00	0.00	43.00
B2g26170	0.71	0.03	0.43	11.00	28.00	1.00	3.00	43.00
B2g26180	0.67	0.00	0.44	14.00	29.00	0.00	0.00	43.00
B2g26300	0.71	0.05	0.46	26.00	10.00	2.00	5.00	43.00
B2g26650	0.87	0.02	0.25	5.00	36.00	1.00	1.00	43.00
B2g26710	0.57	0.02	0.51	18.00	24.00	1.00	0.00	43.00
B2g26780	0.55	0.02	0.52	19.00	23.00	1.00	0.00	43.00
B2g26840	0.85	0.02	0.28	6.00	36.00	1.00	0.00	43.00
B2g26890	0.87	0.02	0.25	5.00	36.00	1.00	1.00	43.00
B2g26990	0.57	0.02	0.51	18.00	24.00	1.00	0.00	43.00
B2g27110	0.88	0.23	0.36	0.00	33.00	10.00	0.00	43.00
B2g27350	0.86	0.00	0.24	36.00	6.00	0.00	1.00	43.00
B2g27400	0.95	0.00	0.09	2.00	41.00	0.00	0.00	43.00
B2g28450	0.98	0.05	0.09	0.00	40.00	2.00	1.00	43.00
B2g30170	0.80	0.00	0.32	32.00	8.00	0.00	3.00	43.00
B2g31720	0.55	0.05	0.54	18.00	22.00	2.00	1.00	43.00
B2g32310	0.50	0.05	0.54	20.00	20.00	2.00	1.00	43.00
B2g32910	0.53	0.08	0.57	17.00	19.00	3.00	4.00	43.00
B2g33060	0.54	0.07	0.56	21.00	18.00	3.00	1.00	43.00
B2g34240	0.92	0.00	0.14	36.00	3.00	0.00	4.00	43.00
B2g34450	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g34550	0.91	0.05	0.21	3.00	38.00	2.00	0.00	43.00
B2g34660	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g34710	0.96	0.02	0.09	1.00	40.00	1.00	1.00	43.00
B2g34820	0.92	0.00	0.15	3.00	35.00	0.00	5.00	43.00
B2g34840	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g34880	0.73	0.07	0.46	29.00	10.00	3.00	1.00	43.00
B2g34970	0.84	0.00	0.27	36.00	7.00	0.00	0.00	43.00
B2g35020	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g35070	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g35150	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g35590	0.84	0.00	0.27	7.00	36.00	0.00	0.00	43.00
B2g35690	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g35720	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g35740	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g35750	0.95	0.00	0.10	2.00	35.00	0.00	6.00	43.00
B2g35800	0.95	0.05	0.13	1.00	40.00	2.00	0.00	43.00
B2g36160	0.68	0.02	0.46	28.00	13.00	1.00	1.00	43.00

B2g36330	0.95	0.00	0.10	2.00	37.00	0.00	4.00	43.00
B2g36400	0.95	0.00	0.10	2.00	37.00	0.00	4.00	43.00
B2g36730	0.95	0.00	0.09	2.00	41.00	0.00	0.00	43.00
B2g36760	0.95	0.00	0.10	2.00	37.00	0.00	4.00	43.00
B2g36820	0.95	0.00	0.10	2.00	37.00	0.00	4.00	43.00
B2g36910	0.84	0.00	0.27	7.00	36.00	0.00	0.00	43.00
B2g36920	0.81	0.03	0.34	6.00	27.00	1.00	9.00	43.00
B2g36930	0.95	0.00	0.10	2.00	37.00	0.00	4.00	43.00
B2g36970	0.95	0.00	0.10	2.00	36.00	0.00	5.00	43.00
B2g37000	0.92	0.00	0.14	3.00	36.00	0.00	4.00	43.00
B2g37750	0.81	0.03	0.34	7.00	31.00	1.00	4.00	43.00
B2g37770	0.83	0.03	0.31	6.00	31.00	1.00	5.00	43.00
B2g37790	0.68	0.00	0.43	13.00	28.00	0.00	2.00	43.00
B2g38440	0.67	0.05	0.48	3.00	36.00	0.00	4.00	43.00
B2g38480	0.92	0.00	0.14	7.00	36.00	0.00	0.00	43.00
B2g39180	0.95	0.09	0.17	2.00	37.00	0.00	4.00	43.00
B2g39280	0.69	0.00	0.43	3.00	35.00	0.00	5.00	43.00
B2g39290	0.97	0.07	0.13	0.00	39.00	3.00	1.00	43.00
B2g39360	0.53	0.00	0.50	0.00	40.00	3.00	0.00	43.00
B2g39420	0.66	0.00	0.45	2.00	37.00	0.00	4.00	43.00
B2g39700	0.83	0.30	0.45	2.00	37.00	0.00	4.00	43.00
B2g39710	0.64	0.03	0.49	3.00	36.00	0.00	4.00	43.00
B2g39790	0.95	0.00	0.10	19.00	23.00	1.00	0.00	43.00
B2g40480	0.79	0.00	0.33	2.00	41.00	0.00	0.00	43.00
B2g40820	0.68	0.02	0.46	1.00	40.00	1.00	1.00	43.00
B2g40930	0.85	0.02	0.28	0.00	39.00	4.00	0.00	43.00
B3g19870	0.56	0.79	0.35	7.00	2.00	34.00	0.00	43.00

Appendix H

Allele frequency distribution of ‘Arina’ and ‘Forno’ alleles in 43 wheat cultivars covering different areas of the world with main focus on Europe



The graphical representation of the ‘Arina’ and ‘Forno’ alleles distribution in 43 genotypes. The horizontal axis shows the names of the SNP markers. The vertical axis shows the percentage of ‘Arina’, ‘Forno’ and intermediate alleles for each SNP marker in 43 wheat genotypes. Yellow bars represent the ‘Arina’ allele, blue bars represent the ‘Forno’ allele, green bars represent the heterozygous allele and grey bars represent missing data. All the tested SNP markers belong to the introgressed region in ‘ArinaLrFor’ between the flanking markers *barc128-gwm131*.

Appendix I

Sequences of the SSR marker, Stk11 and probes used for Forno BAC library screening

Probe name	Forward primer sequence	Reverse primer sequence	Annealing temp
STK11	GATTGAGATCAAACGATGCC	GTCAAACTTCTATAACTTCG	60°
RP3M14	CGCAGTTGTTGGTGCTAAGG	ACGCTTCGGTGTCGTAACCT	55°
Fbox2F/4R	ATCTCCTCCTACTTGCAAAT	TGCTCACCATCTCCCAGTCC	58°
STK1F/2R	GATTGCTGTAAAGCTGCTCCG	ATATCCCTTTGATAATTCCG	55°
TREP15F/18R	TGTTTCCTTCTTTGAGGCGTC	CTGACAACCTTGCCAAACAC	65°
B38460.3	CTTTGCCTTCTCACCTCCAG	GCTTTAGAGGCGTTGTTTCG	60°
CTREP51	ATTGCACTTTCACCATAACC	GTGTTCGGAATAAACCTGAGC	62°
TREP61F/62R	GGAATGTCAGGATCCTGTTTG	CAATCAAGCATGAAGTAGGG	60°

Appendix J

Primer sequences of developed SNP markers designed from SNPs between flow-sorted sequences of 'Arina' and 'Forno' using chromosome 1BS-specific BAC sequence contigs and genotyped by KASPar genotyping

KASP name	Primer Allele Forno	Primer Allele Arina	Primer Common	Allele Forno	Allele Arina
B2g38570	CTATCCCAGCACAGAACAATACAATCAAA	CCCAGCACAGAACAATACAATCAAC	CCCCCATTTGGTAAAAAAGGCTGTAATTT	T	G
B2g38540b	CTATCTGCCCCAGTCTCCGGG	GATCTGCCCCAGTCTCCGGA	GGAGTATGTTCTCTAGTTAATCTTAAGCTA	G	A
B2g38560	CTTAGAGCGCCAAACTGAGCTCC	AATTAGAGCGCCAAACTGAGCTCA	GTCAAAATTCCTTTCATGCACCTTCTCTT	G	T
B2g36430	CTGAAATTTGCACAAGCCCGATGGC	AAGAAATTTGCACAAGCCCGATGGT	TCTCTCTTCATCTTCTCTYTAGTCTTCTT	G	A
B1g21030	CTGCAAAGTGGAACGACCAACTG	CTGCAAAGTGGAACGACCAACTA	GTGTCCAAGGCAAACTGCTGCATTA	G	A
B2g36550	CTATTTGTTGCCTTCCACTATGTGGCT	GTTGCCTTCCACTATGTGGCC	ATGTGGCATACTTTTGCGAAATATGGCAT	A	G
B3g46620	CTATTCAAACCTATCACGCTTCAGGG	CATTCAAACCTATCACGCTTCAGGA	ACCATTAGCATTCAACAATTCAGACACAAA	G	A
B1g60960	CTGCGACCACACTGTGCAGGAA	GCGACCACACTGTGCAGGAT	GGTGCTGATCTTCCCCTGGACAA	T	A
B1g20130a	CTGTCCAAGGCAAACTGCTGCG	GTGTCCAAGGCAAACTGCTGCA	CTTCTTTCTGCAAAGTGGAACGACCAA	G	A
B1g20130b	CTATTGGAATTGGAAAACCTTGTCAAAATTTG	ATATTGGAATTGGAAAACCTTGTCAAAATTTT	CAAATGCATGTCTTCTGGCTAACATCAAT	G	T
B2g44330a	CTGGTATAGGGTGTTGAGCATCTTCAT	GTATAGGGTGTTGAGCATCTTCAC	TACAGTCGGAGCAGGCCCAGAA	T	C
B2g44330b	CTCCGGGCTCAGGGCCTCC	CCGGGCTCAGGGCCTCG	ATCGCCGAACGGGAAGCGGTA	C	G
B4g08060	CTGCATGATTTATACTTGCTATATGACGTTT	GCATGATTTATACTTGCTATATGACGTTC	GTCACTTGACTACACAGCTCTTTGGAA	A	G
B2g20340a	CTAAAACTTGCTTTTAACCAATAAAAACAGACTAT	ACTTGCTTTTAACCAATAAAAACAGACTAC	GCGAGGCTAAGATCCCTTTTTATGGAA	T	C
B2g20340b	CTGTATTATAAGACTGATCTTGTAGACGAAC	GTATTATAAGACTGATCTTGTAGACGAAT	CCTTGTGGACAACCTGCAAAACCAA	G	A
B3g58810a	CTCTTTCATCACTCAGAAATTAAGCAG	CCTCTTTCATCACTCAGAAATTAAGCAT	GCCCGTTCCGTTAAGTTCGGTTATT	G	T
B3g58810b	CTGTCAGAATATCTCCTCCTACTTGCAA	CAGAATATCTCCTCCTACTTGCAG	AAGATGAYGATGATGGCTGCTGCAT	T	C
B3g58810c	CTGAAAGAAATGTGCCCGTTCCGG	GGAAAGAAATGTGCCCGTTCCGT	CAGAATTGAGCAGGGAACCCCAAAT	G	T
B2g38520a	CTCGACATGGTGGTCATGGACTAC	ACGACATGGTGGTCATGGACTAT	GGAGCGGGCAGCCCACGTA	G	A
B2g38520b	CTGCATGAAAATTGTGCCATCGTATCG	AGCATGAAAATTGTGCCATCGTATCA	AGGTCCCAGAGAATTGAAATACAACACTT	C	T
B2g38500	CTCCTTCAGTGCATCATACTCCACC	CCTTCAGTGCATCATACTCCACA	GTCATCAAATCCGTGGGACACAAGTT	G	T
B2g46550	CTCTCCTAGTTCCATCAGAAAAGCG	GTCTCCTAGTTCCATCAGAAAAGCA	CGCACAAATATCATTTAAAGCAAGGAGGTT	C	T
B2g38490	CTCGTTTGGCTTGAAACCAGCATTTGT	GTTTGGCTTGAAACCAGCATTTGC	ATGGTGCCATGWCAATATGGAATGACAT	T	C
B2g38460	CTGGAGATGACAGGAAGTCCGAC	AATGGAGATGACAGGAAGTCCGAA	CAGAACTGTCCAGGAGGCCCAT	G	T
B2g16180a	CTCAGATCCAGGCATTTGATGTGATTG	GCAGATCCAGGCATTTGATGTGATTA	GGCCCGGAGGTGGATAATGTTTATA	G	A
B2g16180b	CTGGATGACAGACGAACATACATC	GCTGGATGACAGACGAACATACATA	CGCCTTATACCTCTCTTTCTTGATATCAA	G	T
B1g75600	CTGAATTCACTTCTCCGTTTTTATATTCTGG	GAATTCACTTCTCCGTTTTTATATTCTGA	CCGCAGAACCTAAGCATGAACAGAT	G	A
B3g30440	CTGCCCTGTCCACTAACATCTCGA	CCCTGTCCACTAACATCTCGG	CAAAAAGGGGCCGAAGTGGAAAGAT	A	G

Acronyms

APR: Adult Plant Resistance

AUDPC: Area Under Disease Progress Curve

BAC: Bacterial Artificial Chromosome

BES: BAC end sequencing

BLAST: Basic local alignment search tool

bp: base pair

CDS: Complementary DNA Sequences

CTAB: Cetyl Trimethyl Ammonium Bromide

FL: Fraction Length

FPC: FingerPrint Contig

HRM: High Resolution Melting

InDel: insertion deletion

ISBP: Insertion Site Based Polymorphism

IT: Infection Type

KASPar: Kompetitive Allele Specific PCR

LTC: Linear Topology Contig

LTN: leaf Tip Necrosis

MTP: Minimum Tiling Path

NGS: next generation sequencing

NIL: Near Isogenic Line

PCR: Polymerase Chain Reaction

QTL: Quantitative Trait Locus

RIL: Recombinant Inbred Line

SNP: Single Nucleotide Polymorphism

SSR: Simple Sequence Repeat

SSRIT: simple sequence repeat identification tool

TDS: target duplication site

TE: transposable element

Acknowledgements

The last five years of my life were a roller-coaster ride. It was a great learning experience and I come out as a much more confident person. These five years of my PhD guided me what I aspire to become and where I want to see myself in the near future. I would like to take the opportunity to thank the important persons in my life who guided me and were with me throughout my journey of five years.

First and foremost, I am grateful to Prof. Dr. Beat Keller for giving me this great opportunity to work in his lab and introducing me to the world of PLANT-PATHOGEN INTERACTIONS. I am highly thankful for his great guidance and his availability for discussions. I am also thankful to him for the freedom he provided to work in my own "Messy" way sometimes in the lab. I enjoyed the nice and friendly working atmosphere in the lab.

I am also thankful to Dr. Simon Krattinger for his great supervision throughout my PhD. I appreciate him for patiently answering my "sometimes very stupid" querries. I am also thankful to him for correcting my thesis in shortest possible time, though I never managed to be as fast as him. Thanks are also due to Dr. Thomas Wicker for introducing me to the fantasy land of BIOINFORMATICS. His sense of humor and joyful nature creates a positive atmosphere in the lab and gives motivation to sit for hours in front of the computer and play with the data. I am also thankful to Prof. Dr. Ueli Grossnikalaus for being a part of my thesis committee and for showing his interest in my research. I am also thankful to the collaboraters, Dr. Abraham Korol, Dr. Tzion Fahima, Dr. Zeev Frenkel, and Dina for providing the 1BS physical information; Dr. Helene Berges and Dr. Arnaud Bellec for the BAC library construction and screening; Dr. Jaroslav Dolezel and Hana Simkova for the flow-sorting of wheat 1B chromosome.

A great deal of thanks to my master student, my lab-mate and my very good friend, Linda Lüthi. She was always there to share my little moments of joys, be it a successful PCR, or successful rust infections in the lab. I owe her a lot for accompanying me during the field seasons and for writing my summary in German. I am also thankful to Ping for being a great lab-mate and always ready to discuss with me. I am thankful to Gabi, the former lab member and Gerry for their technical support during the first three years of my PhD and for being such great friends. I relish discussing the Indian cuisine and sometimes sharing my food with them. I am also thankful to Bea for her help during the field seasons and in handling of plants in the greenhouse. Thanks are also to Kari and Christian for taking care of the plants in the greenhouse. I am also thankful to the librarian Martin Spinnler for his help in arranging the books or papers whenever I needed.

I owe a great deal of thanks to my two partners-in-crime and friends forever, Aditi and Sanae. Without them, I would not have been able to finish my thesis. Their presence always gave me strength, courage and positive energy. I always cherish every single second I spent with them during these years. They both are my soulmates and all the problems tend to resolve so easily whenever they were beside me.

I find myself lucky to be a part of this lab as I make few friends for life. I also want to thank all of them. Lisi for being there in my tough as well as happy times during these years. I cherish the time I spent with her, be it in the lab or outside the lab. She always encouraged me and gave me positive vibes which motivated me to work in the lab. Francis for being a great friend and mentor sometimes during my work. His dry humour and hardworking nature always amazed me and motivates me to work even harder. I am very thankful to Severine who with her warmth and caring nature provided me strength and courage. I am also thankful to Daniel for his support and friendship. I am thankful to my

snicker-giver, Stefan. His incredible comic timing always made me laugh. Thanks are also due to Margarita and Simone for being such great friends and for the time I spent with them outside the lab. Thanks a lot to Jan for all his help in teaching me how to operate Linux and for being so patient with me. I am also thankful to Javi for being a great friend. I am also thankful to Anupriya and Shibu, my small little Indian community who made me feel closer to India.

It is never enough to thank my friends outside the lab, Harsh, Alok, Kumar, Nand, Ravi, Sarvesh and Garima. I am lucky to have them in my life. No matter whatever the situation is, they are always there for me. I relish cooking together in the evening, watching movies and playing games which always took away all my stress. I am grateful to my dad, JaganNath Singla, my mom, Sunita Singla, my brother, Gaurav Singla and my sister-in-law, Anjana Singla for always standing by my side. I am highly thankful to my two nieces Aditi and Anvi who with their innocence always bring smile on my face. Last and most important, I thank the Almighty God for giving me strength and courage to face the difficulties in my life and for giving me this opportunity to thank all of them who contributed in my journey of five years of my PhD.

Curriculum vitae

Family name	SINGLA
First name	Jyoti
Date of birth	30 April, 1984
Place of Origin	Sangrur, Punjab, India

Education

June 2002-June 2006	B.Sc. Agriculture (with Hons. In Plant Breeding and Genetics)
August 2006-December 2008	Masters in Plant Breeding and Genetics, Punjab Agricultural University, Punjab, India
	Master's supervision: Dr. Dharminder Pathak
	Master's thesis title: **Prediction of F_2 performance in *Gossypium arboreum* L. based on genetic diversity of parental lines at molecular level**
May 2010-present	PhD at the Institute of Plant and Molecular Biology, University of Zurich, Switzerland
	Supervision: Prof. Dr. Beat Keller
	Title: **Genetic characterization and high resolution analysis of *Lr75*: a novel adult plant leaf rust resistance gene**